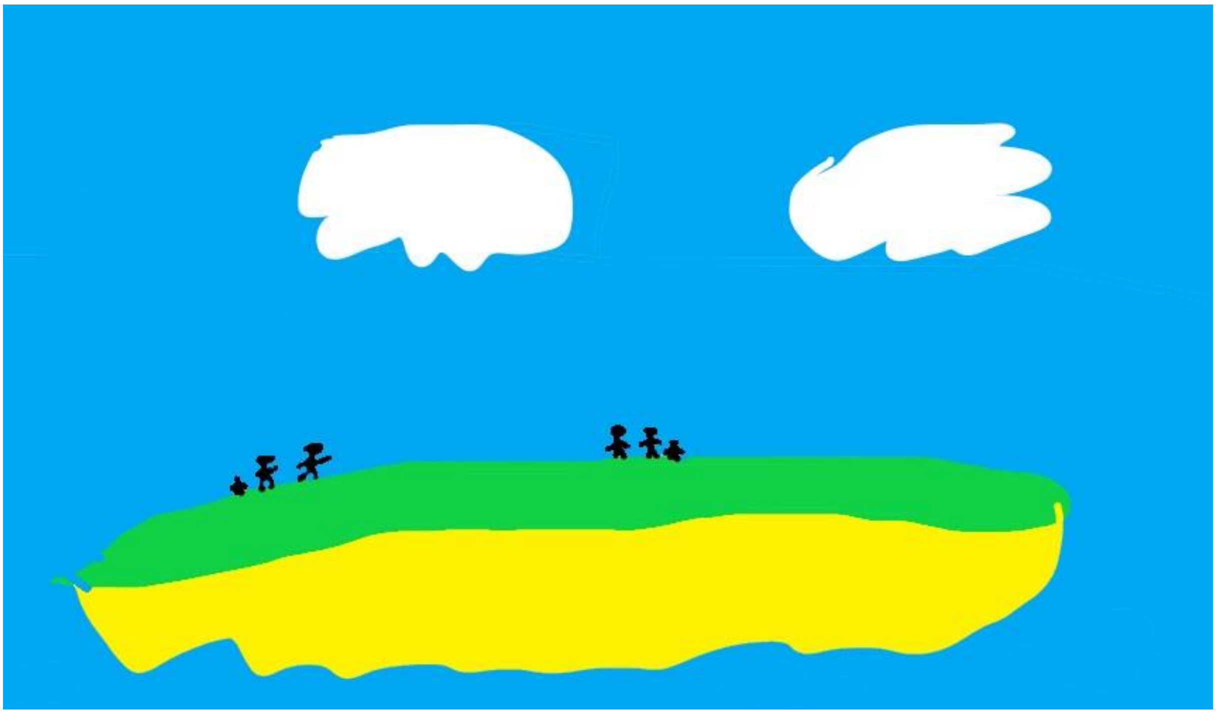
The Prehistory of Language: A Triangulated Y-Chromosome-Based Perspective

Dr. Michael St. Clair, PhD
and the Genetic-Linguistic Interface Project

The Prehistory of Language

A special thanks to Philip for his contribution to this project:



An uncharted island in uncharted linguistic waters.

The Prehistory of Language: A Triangulated Y-Chromosome-Based Perspective

by *Michael St. Clair*

The Genetic-Linguistic Interface

The Genetic-Linguistic Interface Project
Trueffelweg 2
70599 Stuttgart, Germany

E-Mail: mstclair@genlinginterface.com
Project Website: https://genlinginterface.com/

Attribution should include the following information:

St. Clair, Michael R. 2021. *The Prehistory of Language: A Triangulated Y-Chromosome-Based Perspective*. Stuttgart, Germany: The Genetic-Linguistic Interface. https://genlinginterface.com/wp-content/uploads/2021/11/St-Clair-2021.pdf

ISBN (Online PDF Monograph): 978-3-00-071045-2.

Cover image: *Leaving Africa*. © 2021 Michael St. Clair, Stuttgart, Germany, all rights reserved.

Table of Contents

CONTENT PAGE

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

List of Figures

List of Figures

List of Tables

Preface

This monograph represents the end of the Genetic-Linguistic Interface Project that began in 2013. Over the last eight years I posted two previous papers that presented my progress into uncharted linguistic waters, the first in 2016, and the second in early 2020. Now, at the end of 2021, I finally got it right.

I prepared this monograph for my fellow linguists. Nevertheless, I hope other academic researchers will take an interest in the work, especially geneticists, archeologists, anthropologists, and earth scientists. Those that have a general interest in language and genetics are also cordially invited to read my monograph. Please feel free to drop me an email if you have questions or comments.

My path to deciphering the prehistory of language takes advantage of a new research direction that arose roughly 40 years ago in the field of genetics. As the result of sequencing technology, researchers began to utilize molecular genetic variation to explore human evolutionary history. Along the way, some attempted to extend this new research direction even further with the idea that genetic tools can explain the prehistory of language. Genetic and linguistic variation should have a good correlation as we inherit our genes and the mother tongue from our parents. Nevertheless, deciphering language prehistory with genetic data required resolution of several questions. Should we use contemporary DNA, or ancient DNA, or both? Should we use mitochondrial, Y-chromosome, or autosomal markers? Should we build models of language prehistory with statistical methods? Or should we build models with a synthesis of archaeological and paleo-climatological data? With this monograph, I suggest that we employ triangulated Y-chromosome-based modeling as a methodological solution for deciphering the prehistory of language with genetic tools.

My research had identified at least 110 linguistically informative Y-chromosome mutations. The evolutionary history of these mutations suggests that the story of language begins over 100 thousand years ago when *Homo sapiens* migrated out of Africa. Subsequent migrations as well as cultural and evolutionary adaptations then explain the expansion of language to the four corners of the globe. A discussion of this expansion includes Lake Mungo man in Australia, the mammoth steppes of Eurasia, the humid phase of the Sahara Desert, the bidirectional migration of reindeer herders along the Arctic Circle, raised field agriculture along the rivers of the Amazon rain forest, the arrival of rice agriculture in South Asia, malaria in the tropics, and hypoxia on the Tibetan Plateau.

IMPORTANT – PLEASE READ

This monograph avails itself of Supplementary Figures and Tables that can be downloaded from the Genetic-Linguistic Interface website: https://genlinginterface.com. All of the supplementary figures have been combined into a single downloadable PDF document. Similarly, all of the supplementary tables have been combined into a single downloadable PDF document. The “bookmarks” feature of the PDF viewer facilitates navigation of both documents.

In the monograph, supplementary tables and figures are highlighted in blue text to distinguish these resources from tables and figures that appear in the text itself.

The supplementary figures mostly consist of phylogenetic diagrams. These diagrams provide a visual aide that greatly facilitates an understanding of phylogenetic relationships, a topic that appears regularly in the monograph. If I may offer a suggestion, the reader may want to print these diagrams ahead of time or, alternatively, open the PDF document on second computer screen.

Chapter 1: Overview.

Section 1. Introduction.

The correlation between linguistic and genetic diversity is rather straightforward: we inherit our genes and the mother tongue from our parents (e.g., Cavalli-Sforza 2000). Moving beyond the theoretical correlation, however, requires resolution of several methodological questions. Should we use contemporary DNA, or ancient DNA, or both? Should we use mitochondrial, Y-chromosome, or autosomal markers? Should we build models of language prehistory with statistical methods (e.g., Bayesian analysis)? Or should we build models with a synthesis of archaeological and paleo-climatological data? This monograph introduces triangulated Y-chromosome-based modeling as a methodological solution for exploring the prehistory of language with genetic tools.

Section 2. Overview of Research Process.

The research process utilized for this monograph was greatly influenced by Peter Bellwood and his 2005 monograph *First Farmers: The Origins of Agricultural Societies.* He successfully integrates a synthesis of archaeological, climatological, and linguistic perspectives to explain the evolution of early agriculture, a cultural adaptation that arose independently in several region of the world. This work underscores the benefits of utilizing a large dataset drawn from a large cross section of human cultural diversity. Distinct patterns of human cultural evolution surface through the analysis. One striking observation from the Bellwood's monograph is that the contemporary distribution of several language families is linked with the expansion of early agriculture. This observation stands as an important component of his *early farming dispersal hypothesis*, a topic that surfaces repeatedly in this monograph.

As previously mentioned, this monograph explores *triangulated Y-chromosome-based modeling* as a methodology solution for deciphering the prehistory of language. "Y-chromosome based" describes the initial step in the model building process, the identification of informative Y-Chromosome mutations among contemporary populations for which language has a strong ethnic component. The next step in the model building process is the use of "triangulation" to explain why a mutation attains a significant frequency among speakers of a language family. The concept of triangulation is borrowed from the field of navigation and describes a technique that defines your position at a point where three lines converge on a map. Similarly, my research attempts to draw conclusion at a point where several independent lines of evidence converge: the contemporary distribution of Y-chromosome mutations; phylogenetic relationships; language classification; the archaeological record; the paleo-climatological record; ancient Y-chromosome DNA; and other marker perspectives such as mitochondrial DNA.

Triangulated Y-chromosome-based modeling is very much akin to how a public prosecutor would build a criminal case using circumstantial evidence. In a criminal proceeding, the best evidence is provided by eyewitnesses. Well, nobody saw the car accident. Nevertheless, Mr. Smith is suspected of leaving the scene of an accident. According to police logs, the telephone pole was damaged between 0100 and 0300. Forensic analysis of his car suggests that Mr. Smith collided with the object. Finally, the tavern owner reports that Mr. Smith drove home at 0200 and was intoxicated. Similarly, in the field of linguistics, the best data stem from the historical record. Obviously, we cannot investigate the prehistory of language with this data source. A good "circumstantial case" can, nevertheless, be built using Y-chromosome mutations. We inherit language and genes from our parents. As such, the prehistory of language rides the coattails of Y-chromosome mutations and their evolutionary histories.

The data utilized for my research has been extracted from peer-reviewed studies and reports assembled over a fifteen-year period. My analysis of these data is greatly facilitated by the construction of databases. Additionally, the presentation of my research generally follows the nomenclature system adopted by Y-Chromosome Commission in 2002. The term haplogroup describes a unique segment of human Y-chromosome variation and is akin to the volumes of an encyclopedia. This explains why chapter two of my monograph features haplogroup A and chapter eighteen features haplogroup R. Finally, the discussion in the chapters is supported by supplementary figures and data tables. Supplementary materials can be accessed from the Genetic-Linguistic Interface webpage. **https://genlinginterface.com/**

Section 3. The Search for Informative Molecular Markers.

Cross-disciplinary collaboration between geneticists and linguists has long-standing historical precedent. The development of genetic theory began in 1859 with the publication of *On the Origins of Species* by Charles Darwin. He correlates variation in the natural world with reproductive success. In 1863, August Schleicher, a giant in the field of historical linguistics, published an open letter to a professor in Jena, Germany. In the letter Schleicher stressed that linguistics and Darwinian Theory represent complementary methodologies. One idea that surfaced repeatedly was taxonomic relationships, meaning that over time languages and organisms evolve from a common ancestor. This view of language diversity prevails today in historical linguistics which utilizes tools such as the comparative method to demonstrate how several languages diverged from a common ancestral language (e.g., Trask 1996). Another interesting idea from Schleicher's paper is that factors affecting genetic variation also affect linguistic variation. A simple contemporary example would be the Khoi people of southern Africa and the Waorani people of the Amazon rainforest, who possess significant linguistic and genetic differences because of geographic isolation from each other. Portuguese and Spaniards, on the other hand, exhibit far less genetic and linguistic variation because of close geographical proximity.

The term "marker" refers to a section of DNA. Geneticists have found several different polymorphic markers for measuring genetic variation among human populations. The term "polymorphic" means that a section of DNA can vary from one person to the next. Polymorphic protein markers used to assess human genetic variation are generally referred to as "classical markers" in the literature. In 1919, Ludwik and Hanka Hirschenfeld, two researchers at a military hospital, published a study that proposed the use of ABO blood groupings, one of the classical markers, to find patterns of variation among different nationalities and ethnic groups. In 1994, Cavalli-Sforza, Menozzi, and Piazza published the most comprehensive study of human classical marker variation. However, the study conceded (9-10) that another type of marker, mitochondrial DNA (mtDNA), would be a better choice for population studies, but at the time a sufficient number of haplogroups had not yet been discovered.

During the 1980's improved sequencing technology enabled geneticists to focus on the nucleotide bases that form the rungs of the DNA molecular ladder. This development enabled geneticists to analyze population differences at the molecular level, a technique that provides a higher resolution picture of human genetic variation. From this development Brown (1980) identified mitochondrial DNA as a polymorphic molecular marker. About five years later, in 1985, another molecular marker surfaced. Casanova and others identified the non-recombining region of the Y-chromosome as polymorphic markers in human populations. That same year Hill and others (1985) published the first autosomal marker study. From about 1985 onwards, hundreds of reports have been published by geneticists to detail human autosomal, mtDNA and Y-chromosome DNA variation throughout the world. From a macro-perspective, molecular genetic evidence places the origins of *Homo sapiens* in Africa. Cann, Stoneking and Wilson, for example, published a paper in 1987 asserting that female human beings trace their genetic history to a woman living in Africa about 200,000 years ago, the so-called "mitochondrial Eve." African origins for the so-called "Y-chromosome Adam" emerged several years later using data from the non-recombining region of the Y-chromosome (Underhill et al. 2000).

Section 4. Important Key Concepts.

4.1. Overview.

Anthropologists, archaeologists, and geneticists have recognized the potential of Y-chromosome data as a tool for deciphering the human past. Among the linguists, however, the potential of this perspective remains largely unexplored. Accordingly, this section offers several important concepts related to the human Y-chromosome with the goal of persuading linguists that Y-chromosome data are a powerful tool for investigating the prehistory of language. Additionally, the ability to elucidate the prehistory of language with genetic tools represents a recent development. I utilize triangulated Y-chromosome-based modeling. Paleogenomic modeling is an alternative methodology that is under development at the Max Planck Institute in Germany, the David Reich Lab at Harvard University, and elsewhere. Academic researchers may want to consider whether I offer more reliable alternative, a conclusion that can be drawn from this monograph.

4.2. Recombination.

Jobling and Tyler-Smith published an interesting paper in 2003 that paints the Y-chromosome as a non-conforming marker that fails to follow the genetic rules. In order to understand how the Y-chromosome behaves differently from other genetic markers, it is necessary to briefly discuss Mendelian genetics, which is often part of high school and introductory college biology instruction. According to Mendelian genetics we inherit our genes from both parents. However, the Y-chromosome plays by its own genetic rules in that it is only passed from a man to his son. The Y-chromosome is one of the two sex-chromosomes in the human genetic inventory, or human genome. The other sex chromosome is the X-chromosome. During reproduction, two X-chromosomes yield female offspring, and an X-chromosome and a Y-chromosome yield male offspring. Another "rule" of Mendelian genetics is recombination. During human reproduction, the genetic cards are essentially "reshuffled," or more precisely, recombination occurs. This explains, for example, why I have blue eyes and my son has brown. However, the genetic material contained in the Y-chromosome escapes, for the most part, recombination.

In order to explain how the Y-chromosome avoids recombination, it is necessary to briefly discuss the evolutionary history of this chromosome. The sex-determining locus of the Y-chromosome not only codes for male gender in humans, but in all mammals. This section of the Y-chromosome, however, only represents a fraction of its entire length. During the evolutionary history of mammals, about 300 million years, the Y-chromosome has slowly "degenerated" or degraded (e.g., Lahn et al. 2001). When mammals first evolved, the Y-chromosome "behaved normally" in that the entire chromosome recombined with the X chromosome. Now, as the result of slowly evolving structural decay, about 95% of the entire length of the Y-chromosome has been damaged, emerging in what the geneticists call a "non-recombining region." This large non-recombining region means that during reproduction very little genetic exchange occurs between the X and Y chromosome. Consequently, the Y-chromosome has been transmitted largely intact from one human male to the next for the last 300 thousand years.

4.3. Mutation.

The Y-chromosome is unique due to uniparental inheritance and the absence of recombination. Consequently, males inherit a large section of the human genome that remains unaltered when the genetic cards are reshuffled. However, the non-recombining region of the Y-chromosome can and often varies from one Y-chromosome to the next. Geneticists describe this variation as mutation. In population studies examining Y-chromosome variation, one type of a particularly informative mutation is defined by single nucleotide polymorphisms.

In order to better understand the concept of single nucleotide polymorphisms (or "Snips") it is necessary to focus on the molecular structure of deoxyribonucleic acid, or DNA. The molecular "ladder" of DNA has "rails" formed by alternating sugar and phosphate molecules. The "rungs" of this ladder, known as nucleotides, are formed by bonding two molecules having a nitrogenous base which is either adenine and thymine, or guanine and cytosine. Since the non-recombining region of the Y-chromosome has about 60 million molecular "rungs," or base pairs, geneticists have a vast region of genetic information to harvest the evolutionary history of *Homo sapiens*.

The structure of the base pairs that form the Y-chromosome molecular ladder alternate, meaning the nucleotide bases appear in one of four different combinations: adenine/thymine, thymine/adenine, guanine/cytosine, and cytosine/guanine. A single nucleotide polymorphism occurs when one of the rungs of this molecular ladder changes or mutates. Mutations occur, for example, when a nucleotide is substituted for another, or when nucleotides are added or deleted. For example, the R1b-DF27 mutation involves the substitution of a guanine/cytosine bond with an adenine/thymine bond at position 21380200. It should be noted, however, that mutations like this are extremely rare. In fact, some early Y-chromosome studies (e.g., Underhill et al. 2001) initially referred to these Y-chromosome mutations as "unique event polymorphisms" because a base pair only mutates once during human evolution.

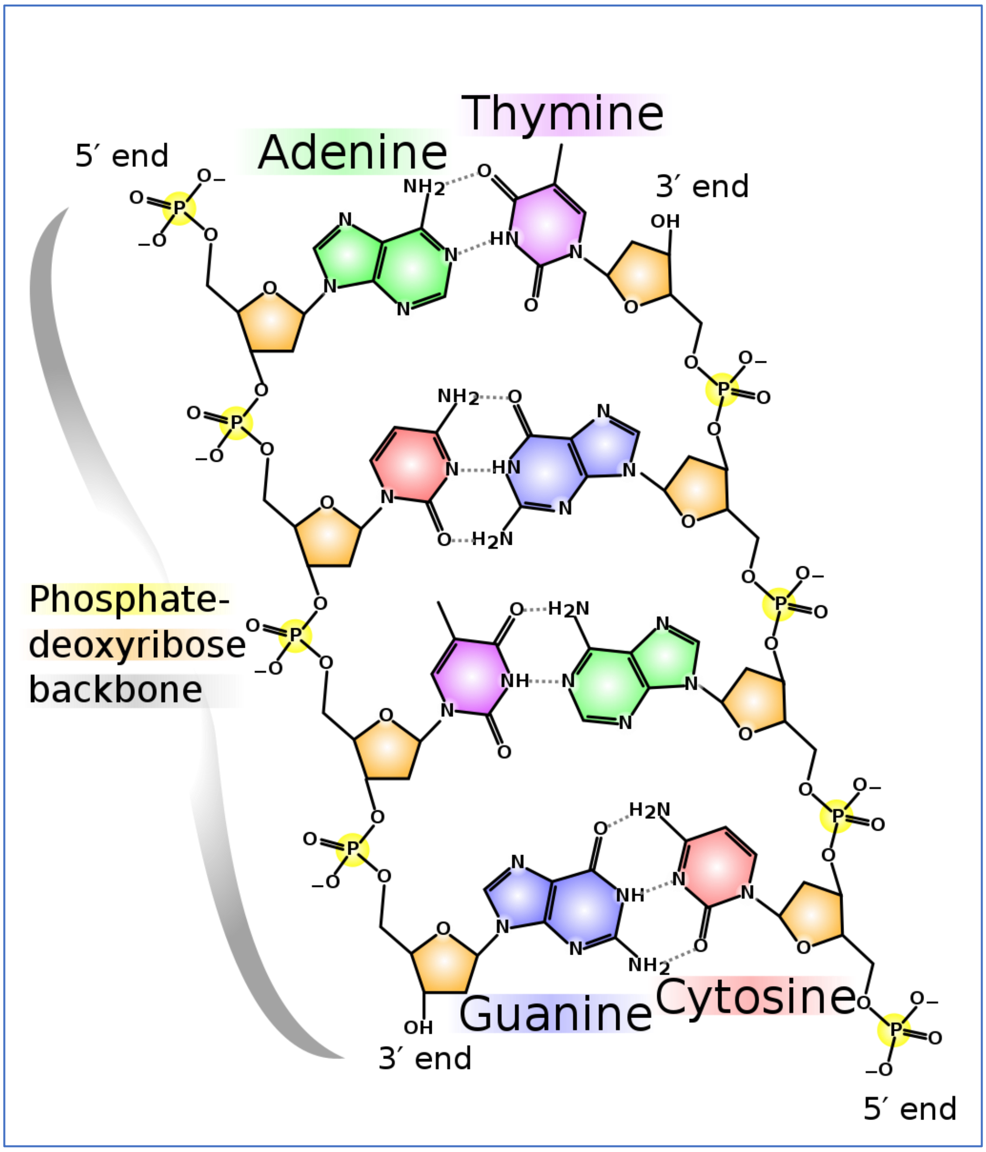
At this point it is important to emphasize a concept known as neutral selection. Those who have taken an introductory biology or physical anthropology course have probably encountered the term "natural selection," initially proposed by the Charles Darwin. This theory accounts for different animal and plant species based on fitness, or survival of the fittest. According to this theory, differentiation among species arose as the result of a mutation that enabled the plant or animal to survive in a given environment long enough to pass on its genes to the next generation. Y-chromosome mutations, however, are classified as selectively neutral, meaning they do not confer any reproductive advantage. Likewise, these mutations are not disadvantageous. Introductory biology courses often emphasize that genetic mutations can be harmful or fatal to living organisms. For example, among humans one of the most recognized harmful genetic mutations is sickle cell anemia. In contrast to sickle cell anemia and other harmful genetic mutations, Y-chromosome mutations are benign. This explains, partially, why Y-chromosome mutations survive while many genetic mutations affect reproductive success and are consequently eliminated from the gene pool.

4.4 Nomenclature and Phylogenetic Relationships.

As explained in the above paragraph, single nucleotide polymorphisms are mutations found within the non-combining region of the Y-chromosome. Geneticists comb the non-combining region of the Y-chromosome to identify these mutations. The presence or absence of mutations, or single nucleotide polymorphisms, can distinguish the genetic history of one population from the next. Y-chromosome single nucleotide polymorphisms are broadly classified as haplogroups, sub-haplogroups, or paragroups. The nomenclature utilized for labeling haplogroups, sub-haplogroups, or paragroups was standardized in 2002 by the Y Chromosome Consortium (YCC 2002). Y-chromosome single nucleotide polymorphism are ordered within a tree-like hierarchical structure. The theoretical Y-Chromosome Adam represents the root of a tree that eventually branches into twenty haplogroups, such as J-M304. The haplogroups, in turn, branch into sub-clades (e.g., J1-M267 and J2-M172). Between Adam and the haplogroups are important mutational steps (e.g., IJ-M429) that YCC 2002 identifies as "paragroups."

Focusing now on the term "haplogroup," this term denotes a major division within the diversity of human Y-chromosome variation. They are akin to the volumes of an encyclopedia. YCC 2002, in their discussion of the standard nomenclature, noted that the label "haplogroup" is "arbitrary." In practice, however, the YCC 2002 nomenclature works surprising well. With the exception of C-M130, the YCC 2002 haplogroups elegantly divide the evolution of contemporary human Y-chromosome diversity. The position occupied by these haplogroup on the global "map" of Y-chromosome diversity is largely determined by contemporary geographic distribution of a haplogroup and its evolutionary history. For example, the Q-M242 haplogroup evolved in Siberia about 30 thousand years ago and represents almost all of the indigenous Y-chromosome variation among Native Americans. The I-M170 haplogroup, on the other hand, evolved in the roughly 40 thousand years, probably in Mediterranean Europe, and now attains a heavy frequency among the contemporary populations of Scandinavia.

Figure 1.1. The Structure of DNA. Source: Wikipedia and Madprime.



The YCC 2002 nomenclature system identifies haplogroups by cladistic name. The cladistic name utilizes set theory and assigns an uppercase letter to identify one of twenty different haplogroups. Then, to identify variants of a haplogroup, the uppercase letter is followed by a combination of numbers and lower-case letters. YCC 2002 also recommended the addition a mutation number to the cladistic name of single nucleotide polymorphism. This number is preceded by a letter such as "M," or "P" or "V." These letters generally identify the laboratory that discovered the mutation. The "M" mutations, such as E-M96, were identified by Peter Underhill, formerly a researcher at Stanford University, who, along with Luca Cavalli-Sforza, played a major role in developing this research tool.

At this point the reader is directed to **Supplementary Figure 1.1** which presents the mutational steps from Y-Chromosome Adam to the main haplogroups. The figure depicts hierarchical relationships among Y-chromosome mutation. Y-chromosome Adam, at the top of the diagram, represents the evolution of *Homo sapiens*. The linguistically informative haplogroups that evolved from Adam are B-M181, D-M174, E-M96, C1-F3393, C2-M217, G-M201, H-M2713, I-M170, J-M304, L-M20, T-M184, M-P256, S-B254, N-M231, O-M175, Q-M242, and R-M207. As previously noted, these haplogroups represent unique segments of contemporary Y-chromosome diversity. Between Y-chromosome Adam and the haplogroups are several important "paragroups" that represent important evolutionary milestones between Adam and the haplogroups. For example, the DR-M168 paragroup represents the genetic ancestor of all the haplogroups that evolved outside of Africa.

As noted above, the initial effort to standardize the Y-chromosome nomenclature was set forth in a 2002 report published by the Y-Chromosome Commission. Since then, important updates for the nomenclature have included Karafet et al. (2008), Oven et al. (2014); Karmin et al. (2015), Karafet et al. (2015); and Poznik et al. (2016). Additionally, since 2002 geneticists have identified thousands of sub-haplogroups (e.g., R1b-DF27) that contribute to our current picture of global Y-chromosome variation. An updated list of these sub-haplogroups is maintained by the International Society for Genetic Genealogy (ISOGG) and their website: **https://isogg.org/**.

Sub-haplogroups provide the most useful mutations for deciphering the prehistory of language. The term "sub-haplogroup" defines diversification of Y-chromosome haplogroups. For example, one very common haplogroup found in Europe is I-M170. The "I" means haplogroup I, and M170 referring to mutation number 170, which was discovered by Peter Underhill at Stanford University (hence "M"). An example of a sub-haplogroup is the I1-M253 mutation, commonly found in Scandinavia. The "1" is used to classify I1-M253 as a sub-haplogroup of haplogroup I-M170. It should be noted that published studies often fail to make a formal distinction between haplogroups and sub-haplogroups. I-M170 and I1-M253 would simply be reported as "haplogroups." Additionally, since the methodology used to build these hierarchical relationships is called cladistics, the terms "clade" and "subclade" are sometime used to label haplogroups and sub-haplogroups. Finally, since 2002 the Y-chromosome tree has grown quite large and the cladistic names of sub-haplogroups have become quite long and cumbersome. For example, the R1b1a1a2a1a2a-DF27 mutation describes a sub-haplogroup that is found on the Iberian Peninsula. Consequently, in practice the nomenclature should be shortened to R-DF27, or perhaps R1b-DF27, when reporting data.

As noted above, the ISOGG website has become the repository of phylogenetic updates for Y-chromosome mutations. However, the focus of this organization is genealogical research, and as such, many of the polymorphisms listed by the organization are not informative markers for linguistic research. Thus, the task for linguists is to identify linguistically informative mutations. For example, the O1b-M95 mutation is a useful marker for deciphering the prehistory of Austro-Asiatic languages. Linguists must keep abreast of the phylogenetic updates posted on the ISOGG website so that old data can be compared to new data. For example, the O3-M122 mutation was re-labeled O2-M122 in 2015.

The reader is now asked to consider several miscellaneous points that facilitate an understanding of the Y-chromosome data and how they are presented:

- **Supplementary Figure 1.1** deviates *slightly* from the YCC 2002 and ISOGG 2017 standard. The F-M89, K-M9, and P-P295 mutations are labeled as higher level paragroups rather than haplogroups. The C-M130 mutation is now a paragroup, and C1-F3393 and C2-M217 become haplogroups.

- I strictly avoid the use of "haplogroup" to describe a "paragroup" or "sub-haplogroup." These terms represent very important distinctions that some genetic studies fail to make.

- Linguistically informative mutations are generally found downstream from the main haplogroups. For example, the A1b-M13 mutation is a useful marker for deciphering the prehistory of Nilo-Saharan languages. To facilitate a discussion of these data, the internal phylogeny of each haplogroup is diagramed by supplementary figures provided for Chapters 2 to 18. These figures are "supplementary" so they can be viewed on a second computer monitor or printed ahead of time.

- Some studies only carry the cladistic identifier (such R1b for the M343 mutation) which makes the task of converting from old nomenclature to new nomenclature extremely difficult. This monograph employs a cladistic identifier with a mutation number (e.g., R1b-M343). Additionally, the cladistic identifier does not extend beyond the second subclade in the monograph. However, the entire identifier appears in the supplementary figures. For example, R1b-DF27 is the short form and R1b1a1a2a1a2a-DF27 is the long form.

- The navigation of hierarchical relationships utilizes the terms "downstream" and "upstream." For example, the R1b-M343 mutation is downstream from the R-M207 haplogroup. Upstream from the R-M207 haplogroup is the P-P295 paragroup.

Linguists should note the phylogenetic mapping of Y-chromosome relationships is akin to mapping linguistic relationships with tree-like language family diagrams. English, for example, is part of the West Germanic sub-branch of the Germanic branch of the Indo-European language family.

4.5. The Dating of Mutations.

The decipherment of hierarchical relationships within the non-recombinant region of the Y-chromosome is aided by the ability to determine, at least roughly, when Y-chromosome mutations evolved. For example, haplogroups D-M174 and E-M96 diverged from the DR-M168 paragroup about 70 thousand years ago (Poznik et al. 2016). Dating methodologies used by geneticists to make these estimates are indeed very complex and involve attempts to determine an average rate of mutation for Y chromosome polymorphisms (e.g., Balanovsky 2017a). Early studies (e.g., Zhivotovsky et al. 2004), attempted to develop a dating methodology by utilizing a type of mutation called short tandem repeats. Today, Karmin et al. (2015) and Poznik et al. (2016) present the latest developments in this ongoing effort. They are able to present more robust estimates by utilizing whole genome sequencing and estimates that are calibrated against the archaeological record and ancient DNA.

4.6. Phylogeography.

The reader is invited to review **Supplementary Figure 1.1** which provides a short summary of where the various Y-chromosome haplogroups are found. For example, haplogroup D-M174 is found in East Asia and haplogroup S-B254 is found in Australia. Furthermore, unlike autosomal markers, it is possible to generate phylogeographic maps of Y chromosome mutations. Such maps trace the frequency of a given mutation across geographical distance. For example, the E1b-M81 mutation exhibits a clinal frequency pattern that increases as one moves westwards across North Africa. The ability to analyze data in this matter, and the ability to estimate when a mutation evolved, helps researchers to explain the prehistoric expansion of people and languages. E1b-M81, for example, suggests that Afro-Asiatic languages and farmers expanded across North Africa during the Neolithic (see Chapter 5).

Prehistoric human expansions often consist of a geographic point of origin and a geographic point of termination. Sometimes the point of origin has the greatest frequency of a particular mutation and over distance the frequency of this mutation diminishes. For example, the J2a-M67 mutation arose in the Near East during the Mesolithic. During the Neolithic this mutation expanded with farmers into Western Europe, and along this route the frequency of the mutation decreased because of admixture with hunter-gatherers already living in the new territory (see Chapter 11). However, some prehistoric migrations show an opposite pattern or cline of haplogroup frequencies, where the point of origin has the lowest frequency of a certain haplogroup, and the terminal end of the migration has the highest. The E1b-M81 mutations, as described above, is an example. The I1-M253 mutation, another example, potentially represents a prehistoric migration from the Pyrenees Mountains to Scandinavia roughly 14 thousand years ago. Along this trajectory the frequency of I1-M253 increases, perhaps because a prehistoric group moved into unoccupied territory, or perhaps they acquired a novel survival strategy that gave them a reproductive advantage (see Chapter 10).

4.7. Population History.

Archaeological and genetic data place human origins in Africa about 300 thousand years ago (see Chapter 2). Around 100 thousand years humans left Africa. Between 60 and 50 thousand years ago humans colonized Europe, South Asia, East Asia, Papua New Guinea, and Australia (see Chapter 4). About 15 thousand years ago people crossed over the Bering land bridge from Asia into North America (see Chapter 17).

From a very clinical point of view the term “population” refers to a group of potentially interbreeding individuals. In practice several factors influence how people choose a partner with whom they eventually have children. These factors may include ethnicity, religion, or socio-economic status. However, the most salient factor for this discussion is geographic distance. For example, southern Africa and the Amazon rainforest are obviously separated by vast geographical distance of several thousand kilometers. Consequently, the Khoisan people of southern Africa and the Waorani people of the Amazon rainforest possess genetic differences because of geographic isolation from each other. From a Y-chromosome perspective, this isolation by distance model explains why the Khoisan of southern Africa have the A1b-V50 mutation and why the Yanomami of South America have the Q1b-M3 mutation; and why the Khoisan do not have Q1b-M3 mutation and why the Yanomami do not have haplogroup A1b-V50 mutation.

The above discussion of the isolation by distance model helps to explain a demographic model called “genetic drift.” The concept posits a leveling of genetic diversity among small isolated populations, a demographic scenario that characterizes most of the human prehistory. *Wikipedia* provides a useful analogy for explaining this concept. A jar is filled with 10 red marbles and 10 blue marbles. Someone is blindfolded and then asked to remove one marble at a time from the jar. At some point the jar will contain either blue marbles or red marbles, but not both. From a Y-chromosome perspective, the populations of the Caucasus provide a good example of genetic drift. Many of the populations practice endogamy, the customs of marrying within the group. This explains the astonishing frequencies of haplogroups G-M201 and J-M304 that are observed in the region (see Chapters 8 and 11).

The term "founder effect" another useful concept that explains the development of genetic differentiation between populations. Founder effect describes a situation whereby mutations frequencies are altered when a group of people separates from a larger population. For example, Austronesian-speakers on Papua Guinea have a mixed ancestry of East Asian and Melanesian Y-chromosome mutations. A sub-population then carried Austronesian languages across Oceania about two thousand years ago. As the result of numerous founder effects, East Asian mutations had disappeared when the expansion terminated at Rapa Nui (or Easter Island). See Chapter 16 for additional information.

The concept of "bottleneck" also helps to explain genetic differentiation among populations. This term describes a situation where perhaps disease or a natural disaster suddenly reduces the size of a population that is isolated and relatively small. This sudden reduction in population reduces the amount of haplogroup variation, and like founder effect, accelerates drift. For example, the Y-chromosome data suggest that the Toba volcano explosion may have produced a bottleneck effect among human populations roughly 70 thousand years (see Chapter 4). Additionally, the data suggest genetic variation in Paleolithic Europe was characterized by the I-M170 and C1-F3393 haplogroups. Today, the only remaining Paleolithic founder mutation among contemporary European is variants of the I-M170 haplogroup. It appears as though the Last Glacial Maximum produced a bottleneck effect in Europe roughly 20 thousand years ago that reduced the size of human populations in Western Europe. Perhaps the bottleneck was caused by a reduction in the number of reindeer and with that, a shortage of food for the hunter-gatherers (see Chapter 18).

Concepts such as isolation by distance, genetic drift, founder effect, and bottleneck represent traditional models of genetic differentiation among populations. While these concepts help to carry a discussion of the Y-chromosome data, they only provide a partial explanation of human Y-chromosome diversity. As noted previously, the non-recombinant region of the Y-chromosome does not alter human reproductive success. Oddly, however, the distribution of Y-chromosome variation is strongly linked to cultural and evolutionary adaptations that have drastically improved the reproductive success of our species. For example, the N1a-M46 Y-chromosome mutation represents an important component that links the domestication of reindeer with the contemporary distribution of Uralic languages. The D1a-P47 mutation, on the other hand, helps to link the origins of Tibeto-Burman languages with evolutionary adaptations that enable people to thrive at high altitudes.

4.8. Advantageous of the Y-Chromosome.

Again, the term "marker" refers to a section of DNA. Common molecular markers used for human population history are autosomal DNA, mitochondrial DNA (mtDNA) and the non-recombining region of the Y chromosome. As such, one potential criticism of this monograph is that it focuses almost exclusively on Y-chromosome variation and excludes autosomal and mtDNA perspectives. The counterargument is that the Y-chromosome provides a perspective of human genetic history that has good-resolution and high-transparency. Mitochondrial DNA data, on the other hand, lack resolution and autosomal markers lack transparency. Indeed, mitochondrial DNA has many of the desirable features of the non-recombinant region of the Y-chromosome, such as the absence of recombination and the ability to order mutations within a phylogenetic tree. However, mtDNA data are gathered from a small section of the human genome that has only 16 thousand base pairs. The non-recombinant region of the Y-chromosome, on the other hand, has 60 million base pairs. As such the Y-chromosome offers a much more resolved picture of human prehistory. For example, mtDNA haplogroups lack counterparts for Y-chromosome R1a-M420 and R1b-M343 mutations as well as the N-M241 haplogroup. The comparison between Y-chromosome and mtDNA data is analogous to the picture quality one obtains from a two-megapixel camera versus a ten-megapixel camera.

Turning now to autosomal studies, one potential benefit of this marker is the ability to obtain a genetic perspective for both genders, whereas the perspective of mitochondrial DNA is arguably for females and the Y-chromosome perspective is arguably for male. However, data from autosomal markers are affected by recombination and thus require incredibly complex statistical analysis. On the other hand, the two uniparental markers, mitochondrial DNA and the Y-chromosome, can be analyzed without complicated statistical methodologies. Moreover, since autosomal data focuses primarily on the frequency of alleles rather than the presence or absence of mutations, autosomal data are not amenable to analysis by means of phylogenetic trees, whereas this is possible with mtDNA and the Y-chromosome (cf. Oven and Kayser 2008; Oven et al. 2014).

As discussed above, one perceived problem with Y-chromosome data is that they only deliver a picture human prehistory for the male gender. In practice, the application of this tool for linguistic research has not confirmed such bias. Rather, this section of the human genome simply acts as an effective "trap" that captures important human demographic milestones that decipher language prehistory for both genders. For example, mtDNA and Y-chromosome perspectives place human origins in Africa (e.g., Oppenheimer 2012).

The huge disadvantage associated with Y-chromosome data (e.g., Jobling and Tyler-Smith 2003) is "ascertainment bias." When investigators collect samples from a population, a small number of samples may skew the actual frequency of a mutation within the population. A larger number of samples, on the other hand, achieves a more realistic picture of genetic diversity. Extending this argument further, more samples yield a more resolved model of language prehistory. Thus, for example, a well resolved picture of language prehistory is available for Indo-European, whereas the picture is highly ambiguous for Eskimo-Aleut and Eyak-Athabaskan (cf. Chapters 11 and 17).

As previously mentioned, the Y-chromosome was identified as a polymorphic marker in 1985. This monograph, which explores the prehistory of language from a Y-chromosome perspective, surfaces almost forty years later. Overcoming the problem of ascertainment bias initially required technological advances that reduced the cost of genetic sequencing. Less expensive sequencing eventually produced a sufficient body of knowledge, which in my opinion, finally occurred at the end of 2020.

4.9. Ancient DNA.

Modern DNA is sequenced from those who were alive when a sample was collected for sequencing. Ancient DNA (aDNA) samples, on the other hand, are taken from the deceased. Generally, the ability to amplify ancient DNA samples diminishes over time. This means, for example, that it is easier to sequence a sample from someone who recently died as opposed to someone who died forty thousand years ago during the Paleolithic. Moreover, environmental conditions play a huge factor that governs how fast DNA degrades over time. For example, DNA from someone buried in the Siberian permafrost remains preserved far longer than DNA from someone buried in the hot and humid tropical jungles of New Guinea. Accordingly, amplification can be a very costly and time-consuming procedure. However, when successful, amplification and radio-carbon dating allow researchers to confirm the presence of mutations at a specific location and time in the past. For example, the Villabruna remains in Italy confirm that Y-chromosome R1b-M343 mutations were part of the genome among the Paleolithic hunter-gatherers of Europe (see Fu et al. 2016).

The effort to harvest ancient DNA was first reported in 1985. Svante Pääbo published a groundbreaking study that reported a novel attempt to sequence DNA from a 2,400-year-old Egyptian mummy. Since then, the use of ancient DNA as a research tool has benefited from efforts to overcome post-mortem decay. To overcome this problem, geneticists have developed sophisticated techniques, such as shotgun sequencing, to amplify and reconstruct damaged sections of ancient DNA samples (see Kivisild 2017 for more details).

Besides post-mortem decay, another technical problem associated with ancient DNA studies is avoidance of contamination, either from modern researchers or microbes. For example, "dinosaur DNA" reported in a study from the 1990's actually came from human DNA (see Cooper and Poinar 2000 for a more detailed discussion). Another example comes from a 2001 study (Adcock et al.) which reported with a great deal of media attention the sequencing a 60-thousand-year-old sample from an ancient Australian, the so-called Lake Mungo Man. They claimed that the mitochondrial DNA of this individual was outside the range of variation for modern humans, and as such, there may have been more than one out-of-Africa migration. The study also asserted that their sample was "authentic" and the researchers had taken the necessary steps to avoid contamination. Fifteen years later Heupink et al. (2016) revisited the 2001 study. The 2016 study determined that the sample utilized by Adcock had been contaminated by European DNA, probably from modern-day researchers.

Several studies have defined protocols for avoiding contamination of ancient DNA samples (e.g., Hofreiter 2001; Pääbo et al. 2004; Gilbert et al. 2005; Knapp et al. 2015; Slatkin et al. 2016). These procedures include how samples are extracted from bone samples, using laboratories dedicated specifically for ancient DNA research, using protective clothing, and decontaminating work surfaces and equipment with bleach or irradiation. Today researchers follow these procedures to avoid contamination and the problem seems to have disappeared.

4.10. Paleogenomics.

Efforts to overcome the effects of postmortem decay and contamination has facilitated a new research direction called "paleogenomics." The main data source for palaeogenomic research is ancient autosomal markers. The first palaeogenomic study appeared in 2010 (Rasmussen et al.) which reported the genome of Saqqaq, a Paleo-Eskimo found in Greenland. Since then, palaeogenomic modeling has evolved into a potential methodology for exploring the prehistory of language. Application of this methodology for investigating the prehistory of language includes a 2015 study by Haak et al. Here, researchers endorsed the so-called Kurgan model of Indo-European language origins based on admixture analysis of 69 ancient DNA samples. Another example comes from a 2019 study (Flegontov et al.) Researchers endorsed the Dene-Yeniseian language hypothesis based on statistical analysis of a small dataset of autosomal markers. Another example comes from a 2020 study (Cui et al.) which endorsed the Transeurasian hypothesis based on statistical analysis of four ancient DNA samples.

The huge problem with palaeogenomic studies is that technical achievement often serves as a license to make huge grandiose conclusions based on very little data. For example, Hofmanova et al. (2016) conclude that Neolithic Aegeans were the source of the Central European Neolithic based on seven ancient DNA samples. Cassidy et al. (2016) suggest Bronze Age steppe nomads are the ancestors of modern-day Irish. This conclusion is based on just four ancient DNA samples. Brandt et al. (2015) assert a large-scale Neolithic invasion of Scandinavia by the Sardinians on analysis one ancient sample. Kivisild (2017) asserts, based on a single ancient DNA sample from Hungary, that Neolithic farmers may have been the source of haplogroup I-M253 variation is Scandinavia. Raghavan et al. (2014) assert that Western Eurasians and Native Americans have a common genetic ancestor based on a single sample. Egfjord et al. (2021) suggest, based on two samples, that Denmark was conquered by steppe nomads during the Bronze Age.

Some anthropologists and archaeologists have enthusiastically endorsed palaeogenomic modeling as a tool for exploring the prehistory of languages (Anthony 2017; Booth 2019; Mallory, Dybob, Balanovsky 2019; Friedlaender and Tucci 2020). An interesting question from their endorsements has arisen. If there is a conflict between ancient DNA and archaeological data, then which data source is more persuasive? For example, is a solid radio-carbon date more reliable than a statistical computation? Perhaps the linguists should ask themselves the same question. Triangulated Y-chromosome-based modeling, an alternate methodology for exploring the prehistory of language, utilizes a non-recombining marker that can be analyzed with archaeological, climatological, and other hard data. Autosomal markers, on the other hand, require statistical analysis. Moreover, the triangulated Y-chromosome-based methodology utilizes contemporary genetic data rather than ancient DNA as the main data source. As such, a significant difference between palaeogenomic and triangulated Y-chromosome-based modeling is their application of ancient DNA for deciphering the prehistory of language. With triangulated Y-chromosome-based modeling, ancient DNA plays a less prominent role in the model building process. They are not the main data source. Rather, they are utilized along with other data sources to interpret the contemporary distribution of genetic mutations.

To illustrate the potential advantages of triangulated Y-chromosome based modeling over the paleogenomic model, it is necessary to revisit the Haak et al. (2015) study. Based on a small dataset of ancient autosomal DNA and statistical analysis, the study links the Indo-European languages of Europe with a massive Bronze Age invasion of steppe nomads from Central Asia. St. Clair (2017) and his triangulated Y-chromosome perspective of Indo-European origins, utilizes a synthesis of Y-chromosome, linguistic, and archeological perspectives. These data demonstrate that the Neolithic triggered a co-expansion of agriculture and language in several regions of the world. This co-expansion explains the contemporary distribution of linguistic "heavyweights" such as Afro-Asiatic, Uralic, Sino-Tibetan, Austro-Asiatic, Niger-Congo, Trans New Guinean, and Austronesian. Taking this a step further, he asks why Indo-European should be an exception to the rule.

Section 5. Chapter Conclusions.

The correlation between linguistic and genetic diversity is rather straightforward: we inherit our genes and the mother tongue from our parents. Far more problematic, in my opinion, is finding a genetic marker that explains this relationship without complicated statistical analysis. The human Y chromosome overcomes this handicap because it is not subject to recombination. This, in turn, facilitates a good-resolution presentation of the human genetic history upon which we can build empirical, transparent, and reliable models of language prehistory. Good resolution stems from the ability of the Y-chromosome to deliver 110 linguistically informative mutations that elucidate the prehistory of language. Empirical and transparent means that Y-chromosome data are amenable to triangulation with other data sources, such as the archaeological record, the climatological record, and language variation. *Triangulated Y-chromosome-based modeling*, in turn, produces far more reliable models of language prehistory because it is built from a convergence of several independent lines of evidence.

Chapter 2: Haplogroup A.

Section 1. Contemporary Distribution of Haplogroup A.

The reader is asked to locate the A1b-V50, A1b-M51, and A1b-M13 mutations in **Supplementary Figure 2.1**, which diagrams the phylogenic relationships within haplogroup A. These markers represent almost all the published data for the haplogroup. The A1b-V50 and A1b-M51 mutations are concentrated in southwestern Africa. A1b-M13, on the other hand, is concentrated in eastern Africa (see, also, **Supplementary Tables 2.1**, **2.2**, and **2.3**. Turning now to linguistic diversity in Africa, A1b-V50 and A1b-M51 are found among populations that speak languages from the Khoisan macro-language family. The A1b-M13 mutation, on the other hand, helps to decipher the prehistory of the Nilo-Saharan, Niger-Congo, and Afro-Asiatic language families.

Section 2. The Evolutionary History of Haplogroup A.

Unlike the other main haplogroups, haplogroup A is defined by a man rather than a mutation. He is called Y-chromosome Adam. A 2013 report (Mendez et al.) suggests that Y-chromosome Adam evolved among a population of *Homo sapiens* who lived in west-central Africa about 338 thousand years ago. This conclusion stems from genetic data collected from an African American having the oldest known haplogroup A lineage, one that is defined by A00-AF6/L1284 mutation. Additionally, the study analyzed variants of this mutation among the Mbo people of western Cameroon. It should be emphasized that when "Adam" was born, other Y-chromosome mutations may well have been present among the humans of northwestern Africa. His Y-chromosome lineage managed, however, to survive 300 thousand years of drought, famine, disease, and violence. Other lineages, which may have been present among his contemporaries, ultimately perished.

The evolution of Y-chromosome Adam in northwestern Africa is supported by the archeological record. Hublin et al. (2017) report the oldest currently known fossil remains of *Homo sapiens*. These remains were found at Jebel Irhoud in Morocco and date to roughly 300 thousand years ago. The report by Hublin et al. is especially significant because previous fossil evidence (e.g., White et al. 2003) had placed human origins in eastern Africa about 200 thousand years ago. As such, the genetic data are now more consistent with the fossil record.

At this point the reader is directed to the top of **Supplementary Figure 2.1** which provides a phylogenetic overview of important mutations within haplogroup A. According to Poznik et al. (2016), A0-V148 and A1-V168 split from A00-AF6/L1284 about 190 thousand years ago, and A1a-M31 and BR-M42 separated from A1-V168 about 160 thousand years ago. As noted previously, A1b-V50, A1b-M51 and A1b-M13 represent almost all of the published haplogroup A frequency data. Dating estimates provided by D'Atanasio et al. (2018) for the phylogenetically equivalent A1b-M32 mutation suggest that A1b-V50 evolved roughly 100 thousand years ago. Naidoo et al. (2020) suggest that A1b-M51 and A1b-M13 evolved roughly 50 thousand years.

Section 3. Early Cognitive Evidence for the Evolution of Language.

The A1b-V50, A1b-M51 and A1b-M13 mutations stand as genetic relics of the African Middle

Stone Age, which began roughly 300 thousand years ago, and ended roughly 20 thousand years ago. Lombard (2012) provides a useful overview of this period of human prehistory in sub-Saharan Africa. Archaeological research of the African Middle Stone attempts to assess the emergence of cognitive and behavioral complexity that is characteristic of *Homo sapiens*. Evidence often includes technological developments such as tools or adhesives. Additionally, evidence of symbolism and ritual may also support the emergence of cognitive and behavioral complexity.

Wilkins et al. (2020) recently published a study that present evidence of human occupation of Ga-Mohana Hill North Rock Shelter in South Africa about 105 thousand years ago. This site is located about 665 kilometers from nearest coastline. This site dispels previously held opinions in archaeology that limited Middle Stone Age human activities to the coastal regions of Africa. Moreover, calcite crystals were found at the site. Those who occupied the rock shelter had collected these non-utilitarian objects from another location. Wilkins et al. (2020) suggest that the crystals may have had spiritual significance for these people. This evidence of symbolism, in turn, may provide support of the emergence of cognitive and behavioral complexity that is characteristically human.

Evidence of mortuary rituals also point to the emergence of cognitive and behavioral complexity. All that lives will die, and we humans ask what happens afterwards. Martinon-Torres et al. (2021) recently published a study that reports the earliest known human burial in Africa. Evidence comes from Panga ya Saida archaeological site along the coast of Kenya. Researchers unearthed the partial skeletal remains of a three-year-old child whom they call Mtoto, which means "child" in Swahili. According to the researchers, the child died about 78 thousand years ago. Those who buried Mtoto wrapped him or her in some type of clothing. The body was then positioned in the grave in a manner that is indicative of a mortuary ritual. This, in turn, suggests a strong emotional attachment with the deceased child and perhaps a belief in the afterlife.

Section 4. Contemporary Linguistic Diversity in Africa.

Language is a manifestation of human cognitive and behavioral complexity that emerged during the Middle Stone Age in Africa. Contemporary linguistic diversity on the continent is discussed in this present chapter as well as in Chapters 3, 5, 11, 16, and 18. As show by Figure 2.1 (below), this diversity consists of the Khoisan macro-language family as well as the Niger-Congo, Nilo-Saharan, Afro-Asiatic, and Austronesian language families. Khoisan, Niger-Congo, and Nilo-Saharan represent languages that evolved in Africa. Afro-Asiatic and Austronesian, on the other hand, evolved elsewhere and then expanded onto the African continent.

Section 5. Southern African Khoisan.

Haplogroup A evolved in Africa about 300 thousand years ago. Today the haplogroup is found almost exclusively among populations living on the African continent. Two downstream variants of haplogroup A, the A1b-V50 and A1b-M51 mutations, represent the genetic relics of pre-agricultural populations in southern Africa whose contemporary representatives are the so-called Khoisan people (e.g., Rosa et al. 2007; Batini et al. 2011). The reader is now directed to **Supplementary Tables 2.1** and 2.2. Among the Khoisan, the A1b-V50 mutation attains a frequency up to fifty percent. A similar percentage is reported for the A1b-M51 mutation.

It should also be emphasized that the term "Khoisan" has both cultural and linguistic components. As noted above, from a cultural perspective the Khoisan are the descendants of pre-agricultural hunter-gatherers who evolved and remained on the African continent. Furthermore, they remained hunter-gatherers and resisted assimilation with Bantu farmers who migrated from west-central Africa to South Africa beginning about five thousand years ago (see Chapter 5 for additional information). Turning now to the linguistic component of the term "Khoisan," these populations speak languages that fall into one of three language families: Khoe-Kwadi, Kx'a, and Tuu. This is quite significant from a linguistic perspective as the Khoisan not only resisted the adoption of the Bantu farming culture, but also resisted shifting to the Niger-Congo languages spoken by the Bantus.

Figure 2.1. Linguistic Map of Africa. Source: Wikipedia and Mark Dingemanse.

The above explanation of the term "Khoisan" facilitates an important discussion of two important points that linguists need to know. First, geneticists use the term "Khoisan" ubiquitously in published reports that describe genetic variation in Africa (e.g., Underhill et al. 2000; Tishkoff et al. 2007; Barbieri et al. 2016). Secondly, as noted by Mitchell (2010), the term "Khoisan" seems to lump too many groups into one basket. Each group actually has cultural, linguistic, and genetic histories that should be evaluated independently.

Section 6. Nilo-Saharan, Afro-Asiatic, and Niger-Congo.

As noted in the previous section, populations having a significant frequency of haplogroup A are concentrated either in southwestern Africa or eastern Africa. Populations in southwestern Africa have the A1b-V50 and A1b-M51 mutations. However, eastern African populations have the A1b-M13 mutation. This mutation attains an especially high frequency (fifty percent or greater) among the Nilo-Saharan populations in this region. Besides Nilo-Saharan speaking populations, the A1b-M13 mutation also attains a significant frequency among some Afro-Asiatic speaking populations in east Africa, such as Amhara, Oromo and Welayta. Finally, the A-M13 mutation has a moderate frequency among some Niger-Congo speaking populations (see **Supplementary Table 2.3** for additional details).

The Sahel region of Africa is a transition zone that runs west to east across Africa, between the Sahara Desert and the Central African rainforest. What is particularly interesting about the A1b-M13 data it that they exhibit a clinal pattern across the length of the Sahel, with the highest frequency in eastern Africa. This observation supports the position that places the geographic origins of Nilo-Saharan languages within the region. Further support for this position stems from a 2010 study published by Gomes and others. Here, researchers analyzed A1b-M13 data and based on their analysis, estimated that Nilo-Saharan populations arose about 15 thousand years ago. Additionally, the same study determined that within the Nilo-Saharan language family, the Eastern Nilotic and Western Nilotic branches diverged and expanded four to six thousand years ago. These dating estimates are significant for defining the Sahel as the geographic point of origin for Nilo-Saharan as they correlate well with climate change that occurred within the Sahel during the early Holocene between 10,500 and 7,300 years ago.

During the Holocene, which began about 12 thousand years ago, warmer temperatures caused the ice glaciers to retreat across the Northern Hemisphere. Further south, the onset of the Holocene triggered a temporary but dramatic change within the Saharan Desert of northern Africa that lasted three thousand years. Normally this area stands as one of the most inhospitable regions of the world, with endless miles of sand dunes, temperatures that approach 50 degrees centigrade, and no rainfall.

However, about 10 thousand years ago, as the result of global climate change, monsoon rains came to the region. During this "humid phase" rain transformed the desert into a savanna ecosystem characterized by grassland and widely spaced trees. Furthermore, the monsoon rain produced rivers and lakes that attracted hunter-gatherers to the region.

Figure 2.2. Bushmen Hunter. Source: Wikipedia and Andy Maano.

A fascinating study by Drake et al. (2011) presents the results of satellite imagery. Their data confirmed the presence of a complex system of rivers and lakes that arose during the early Holocene in the Sahara Desert, something that researchers had long suspected as the result of the archaeological record. The study also discusses artifacts from the humid phase, numerous barbed bone points left behind by the hunter-gatherers who once harvested to food resources within this complex system of rivers and lakes. During the last humid phase in the present-day Sahara Desert, Stone Age people utilized these bone points to make harpoons. This technological adaptation enabled them to harvest the hippos, crocodiles, and fish that thrived here. Drake and others provide a map that illustrates the distribution of these points, which includes the entire Sahel. The study then suggests that the location of these artifacts offers a good correlation with the distribution of Nilo-Saharan languages. To support this position Drake and others present a list of cognates for "crocodile" and "hippo," as found in several of the Nilo-Saharan languages.

Kuper and Kröpelin in their 2006 study focus on the eastern Sahara and effects of climate transformation that occurred during the last humid phase in this region. The eastern Sahara encompasses the Western Desert of Egypt, northwestern Sudan, and parts of Libya and Chad. Interestingly, during the humid phase the Nile Valley was not occupied by people as this region was too "marshy." According to the study, west of the Nile, in the savanna that existed seven to ten thousand years ago, people initially survived by hunting and gathering. People in the region later adopted pastoralism, the herding goats and sheep that came from the Middle East, and cattle that may have their origin in Africa. Then around 7,300 years ago the rain suddenly ended, and region became once again, almost overnight, a desert. As the result of desertification, some populations migrated into the Nile Valley, where the region had become, in the meantime, more habitable as the result of dryer climate. Those that settled along the Nile River later adopted cereal cultivation and became the founding population of Pharaonic Egypt. Other populations, instead of settling along the Nile, escaped desertification of the Sahara by migrating into the Sudan and later eastern Africa.

The above discussion of the 2006 study by Kuper and Kröpelin supports the idea that the current distribution of Nilo-Saharan languages may well be a product of climate change that occurred in the Saharan Desert about 7,000 years ago, when this area reverted from a savannah ecosystem to a dry desert. The origins of this language family seem to be linked to the origins of East African cattle pastoralism. This subsistence strategy stands as the traditional food economy documented among many of the Nilo-Saharan speaking populations, such as the Dinka and Maasai.

The A1b-M13 data may also support the position that Niger-Congo and Nilo-Saharan languages evolved from a common ancestral proto-language. This topic was explored by the researcher Roger Blench in an unpublished 2006 paper with the title *The Niger-Saharan Macrophylum*. Alternatively, the data simply supports the independent evolution of Niger-Congo and Nilo-Saharan languages at a time that coincidently occurs after the desertification of the northern Africa. Turning now to A1b-M13 and Afro-Asiatic languages, the data may support the position that some Nilo-Saharan populations shifted to Afro-Asiatic after the arrival of Southwest Asian agricultural package. This discussion continues in Chapter 5.

Section 7. Conclusions for Haplogroup A.

The origins of Haplogroup A extend back to Y-chromosome Adam. The A1b-V50 and A1b-M51 mutations elucidate the prehistory of the so-called Khoisan language families. The Khoe-Kwadi, Kx'a and Tuu language families are linguistic relics of the African Middle Stone Age and the emergence of modern human behavior and cognition. The A1b-M13 data, on the other hand, help to decipher the prehistory of the Nilo-Saharan, Niger-Congo, and Afro-Asiatic language families. This mutation, along with archaeological and climatological data, suggest that Nilo-Saharan and Niger-Congo may have evolved in the Sahel at the end of the last humid phase. Finally, the A1b-M13 data suggest that the Afro-Asiatic languages of Africa are a Neolithic "import" from Southwest Asia (see Chapters 5 and Chapter 11 for more details).

Chapter 3: Haplogroup B-M60.

Section 1. The Evolutionary History of the B-M60 Mutation.

According to Poznik et al. (2016) the B-M60 haplogroup evolved about 100 thousand years ago. Like haplogroup A (see Chapter 2), haplogroup B-M60 evolved and remained in Africa. The reader is now directed to **Supplementary Figure 3.1** which provides a phylogenic overview of this haplogroup and its informative downstream variants, the B1a-M150 and B1b-M112 mutations. Both markers evolved about 50 thousand years ago (see Barbieri et al. 2016).

Section 2. Contemporary Distribution of B-M60 Mutations.

2.1. Overview.

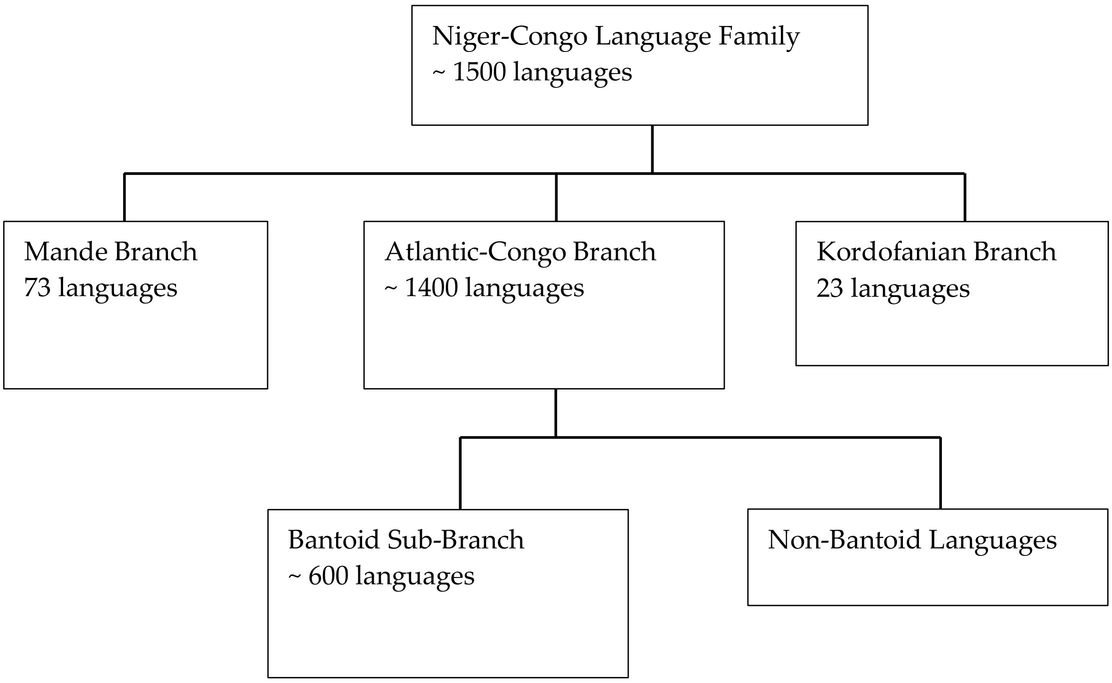
Linguistic variation and traditional subsistence strategies within Africa help to deliver a presentation of the haplogroup B-M60 data. Consequently, this section presents data for the following groups: non-Bantoid farmers, Bantoid farmers, Pygmy hunter-gatherers, Khoisan hunter-gatherers, Hadza hunter-gatherers, Sandawe hunter-gatherers, Nilo-Saharan farmers, and Afro-Asiatic farmers. Focusing now on linguistic variation in Africa, the reader is directed to the language map of Africa provided in Chapter 2 and Figure 2.1. As shown by the map, linguistic diversity in Africa includes the Afro-Asiatic, Nilo-Saharan, and Niger-Congo language families, as well as the Khoisan macro-family of languages. Linguistic diversity in Africa also includes Hadza and Sandawe, two language isolates that are found in Tanzania. They, along with Khoisan, have click consonants, a unique speech sound within the phonemic inventory.

Turning now to traditional subsistence strategies within Africa, these consist of pastoralism, hunting and gathering, or sedentary cereal agriculture. The herding of goats and sheep represent an important cultural relic of Afro-Asiatic languages in northern Africa. Cattle pastoralism is practiced among many Nilo-Saharan speaking populations in eastern Africa. Among the Hadza and Sandawe of Tanzania, the Khoisan of southern Africa, and the Pygmies of the central African rainforest, hunting and gathering stands as the traditional subsistence strategy. Finally, the origins and expansion of Niger-Congo languages throughout sub-Saharan Africa evolved from the cultivation of millet and sorghum.

2.2. Non-Bantoid Farmers.

Within the Niger-Congo language family a discussion of haplogroup B-M60 data is facilitated by a distinction between Bantoid and non-Bantoid languages. The Niger-Congo family contains approximately 1,500 languages that are separated into three main branches: Mande, Kordofanian and Atlantic. According to *Ethnologue* (2016), the 73 languages of the Mande branch are concentrated in central western Africa. The same source places the 23 Kordofanian language in southern Sudan, and as such, they clearly occupy an "outlier" position within geographic distribution the Niger-Congo language family. The remaining Niger-Congo languages (around 1,400 according to *Ethnologue*) fall within the Atlantic-Congo branch. These languages extend from Nigeria to South Africa. Nested deep within the numerous and complex sub-branches of Atlantic-Congo are about 600 languages classified by *Ethnologue* (2016) as "Bantoid." The remaining 800 Atlantic-Congo languages are reported in this discussion as "non-Bantoid.

Figure 3.1. Simplified Overview of the Niger-Congo Language Family. Source: Ethnologue 2018.



Focusing now on non-Bantoid farmers, such as the Yoruba of Benin and the Fali of Cameroon, the presence of the B2a-M150 mutation among these populations represents a significant observation. This support linguistic and anthropological perspectives that identify West-Central Africa as the putative homeland of Niger-Congo languages. Taking this a step further, the B2a-M150 mutation was present in West-Central African among Atlantic-Congo speaker prior to the Bantu expansion. See, also, Scozzari et al. (2012) and Section 3 (below) for more details. Turning now to the B2b-M112 marker, this mutation fails to attain a significant frequency among non-Bantoid speakers. For more details, see **Supplementary Tables 3.1 and 3.2**.

2.3. Bantu Farmers.

As shown by **Supplementary Table 3.1**, the B2a-M150 mutation is clearly a significant marker among the Bantu. Moreover, it is found along the entire geographic range of Bantoid languages, from the Ngumba people of Cameroon to the Zulu of South Africa. The B2b-M112 mutation, on the other hand, is not a significant marker among the Bantu farmers (see **Supplementary Table 3.2**).

2.4. Pygmy Hunter-Gatherers.

As shown by **Supplementary Table 3.2**, the B2b-M112 mutation attains a significant frequency among Pygmy hunter-gatherers. Among Baka populations in Gabon and Cameroon, for example, the reported frequency is 60 percent or greater. A similar figure is attained among the Mbuti of the Congo region. A significant frequency of the B2b-M112 mutation is also found among the Aka of the Central African Republic and the Gyele of Cameroon. Focusing once again on **Supplementary Table 3.1**, the B2a-M150 mutation does not appear to be a significant marker among the Pygmies.

2.5. Khoisan Hunter-Gatherers.

As noted previously in Chapter 2, the Khoisan are the descendants of pre-agricultural hunter-gatherers who evolved and remained on the African continent. The B2a-M112 mutation attains a significant frequency among the Khoisan (see **Supplementary Table 3.2**). The B2a-M150 mutation, on the other hand, does not appear to be a significant marker for the Khoisan except for ||Gana people of Botswana (see **Supplementary Table 3.1**). Among the ||Gana. the mutation attains a perplexing frequency of 79 percent. This presents a topic for future investigation. Perhaps like the Damara of Namibia, the ||Gana were a Bantu group that switched to a Khoisan language (see Rocha and Fehn 2016). Among the other Khoisan populations, for whom data are available, the B2a-M150 attains a low to moderate frequency. A potential source of these mutations among these Khoisan groups is geneflow from Bantu males who became part of the Khoisan groups.

2.6. Hadza and Sandawe Hunter-Gatherers.

The Hadza and the Sandawe people of Tanzania, as noted earlier, are counted among the African populations that speak an isolate language. Among both populations the B2b-M112 mutation also attains a significant frequency, present in about half the Hadza and a third of the Sandawe (see **Supplementary Table 3.2**). However, the B2a-M150 mutation does not attain a significant frequency among both populations (see **Supplementary Table 3.1**).

2.7. Nilo-Saharan Farmers.

Surprisingly, the B2a-M150 mutation attains a significant frequency among some Nilo-Saharan populations: 50 percent of Alur (Congo region), 22 percent of Luo (Kenya), 17 percent of the Ng'arkarimojong in Uganda (see **Supplementary Table 3.1**). The B2b-M112 mutation, on the other hand, fails to attain a significant frequency among Nilo-Saharan populations (see **Supplementary Table 3.2**).

It should be noted that resequencing the B-M60 data obtained by Hassan et al (2008) may yield more B2a-M150 or B2b-M112 data for Nilo-Saharan populations.

2.8. Afro-Asiatic Farmers.

With the possible exception of Cushitic speakers in Tanzania, the B2b-M112 mutation does not represent a significant marker for the Afro-Asiatic populations of Africa (see **Supplementary Table 3.2**). Turning now to B2a-M150, this mutation is virtually absent among Afro-Asiatic-speaking populations (see **Supplementary Table 3.1**).

2.9. Ascertainment Bias.

It should be noted that ascertainment bias might also be a problem with currently available Y-chromosome data for Africa. The data reported in **Supplementary Tables 3.1** and **3.2** are extrapolated from populations for which, in many cases, a relatively small number of samples were collected for sequencing. As such, the reported mutation frequencies may not reflect the actual frequency found within the group. The scarcity of African data stands in sharp contrast to other regions of the world, especially Europe or East Asia, where hundreds of samples from a single population are sequenced to generate data. Finally, it should be emphasized that Africa possesses enormous cultural and linguistic diversity. **Supplementary Tables 3.1 and 3.2** only capture a miniscule amount of this diversity. Hopefully, the future will bring more data for the continent.

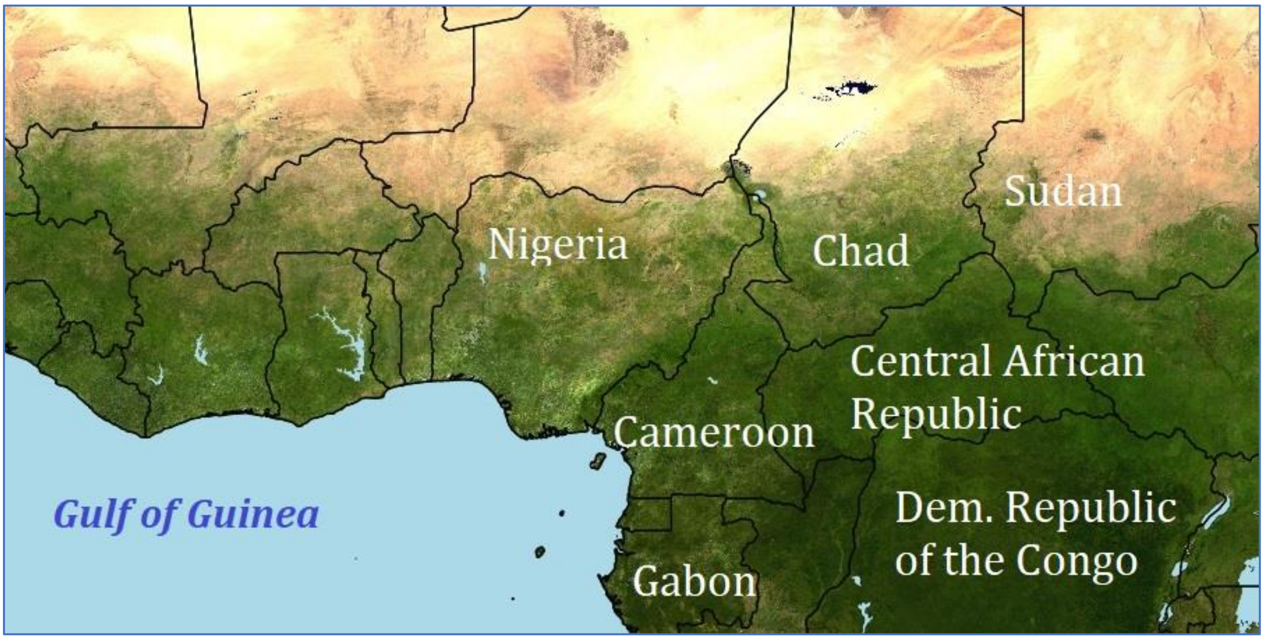
Section 3. The Bantu Expansion and Language Shift among the Pygmies.

Obviously, more genetic data from African populations are needed. Despite this handicap, the available haplogroup B-M60 data identify the B2b-M112 mutation as a genetic signature of hunter-gatherer populations in Africa. The B2a-M150 mutation, on the other hand, stands as a genetic relic of the Bantu expansion through the central African rainforest (for more details, see Berniell-Lee 2009; Batini et al. 2011).

The B2a-M150 and B2b-M112 mutations help to assess male geneflow between Bantu farmers and the hunter-gather populations of Africa. These data identify factors that contributed to a massive language shift among the Pygmies. It is important to note that Pygmy hunter-gatherers inhabited the central African rainforest long before the arrival of farmers. When the agriculture finally expanded into this region, the Pygmies languages became extinct because they shifted to the language of the farmers with whom they had established contact. In most cases they shifted to a Bantoid language. However, the Mbuti people speak Nilo-Saharan languages, which presumably came from an independent incursion into the central African rainforest by Nilo-Saharan-speakers from the Sudan. Similarly, the Baka speak a non-Bantoid language from the Niger-Congo family that presumably came from an incursion of Ubangi-speaking farmers into the rainforest.

Focusing now on the Bantus, the homeland of the Niger-Congo language family straddles the present-day border of Nigeria and Cameroon (Grollemund et al. 2015). As illustrated by Figure 3.2 (below), an ecological transition begins at this location. The savanna of West-Central Africa eventually becomes the rainforest of central Africa. The southward expansion of the Bantus into the rainforest is linked to climatic changes that occurred roughly four thousand years ago. According to Maley et al. (2018), a sudden cooling of the sea-surface temperature occurred in the Gulf of Guinea. This altered the pattern of the monsoon rains and savannah began to appear along the northern periphery of the rainforest. According to Bostoen et al. (2015), the appearance of savannah along the northern periphery enticed people from the Niger-Congo homeland to migrate southwards. As detailed in the same study, around 2,500 years ago, savannah appeared in areas of the rainforest itself. This facilitated travel into the region. Inside the region, the Bantus were able to navigate through the rainforest along the extensive network of rivers in the area. Additionally, Bantu farmers were able to utilize the areas of savannah within the forest to cultivate cereal crops. The primary African cereals cultivated by these farmers consisted of pearl millet, finger millet, and sorghum (e.g., Crowther et al. 2017).

Figure 3.2. Biodiversity of West-Central Africa. Note: Image depicts the eco-diversity and transitions from the Sahara Desert in the north to rainforest in the south.



As noted previously, the Pygmies foraged in the Central African rainforest long before the arrival of the Bantu farmers. Serge Bahuchet (2012) presents a useful anthropological study of Pygmies and Bantu farmers. The term "Pygmy" describes 20 different groups, such as the Aka, Baka and Mbuti. They have an unusual physical characteristic (or phenotype) characterized by short stature. According to some oral traditions, when the Bantu farmers eventually penetrated the rainforest (about three thousand years ago), the Pygmies initially guided them through the "forest world." After the farmers had cultivated areas of the rainforest, they often traded with nomadic or semi-nomadic Pygmy groups in order other to exploit a "common ecosystem."

In his 2012 report, Bahuchet also examines the linguistic distance between contemporary Pygmy groups and their closest farming neighbors. This explains how the Pygmies shifted languages. In some cases, the linguistic distance is close, which in turn indicates that both groups have lived alongside each other for a considerable period. Thus, intense contact over a prolonged period seems to explain language shift among some Pygmy groups. However, in other cases the linguistic distance between contemporary farmer and Pygmy groups is large even though both groups live alongside each other. Furthermore, contemporary exchange between these neighboring Pygmy and Bantu farmer groups is often facilitated by bilingualism rather than language shift on the part of the Pygmies. To explain this observation, Bahuchet provides examples of linguistically close Pygmy and Bantu farmer groups that are now separated considerable geographic distance, sometimes several hundred kilometers. According to Bahuchet, this indicates that sometime in the past both groups co-migrated, perhaps along a river, and then separated. Thus, while the contact may have been intense, the duration of contact may have been relatively short.

As already noted, the Pygmies shifted to the language of Bantu groups that they had encountered. Often language shift is preceded by a period of intense contact between two groups where the social standing of one group is perceived as more prestigious than that of the other group. The Y-chromosome evidence (**Supplementary Tables 3.1** and **3.2**) suggests that male geneflow between Bantu farmers and Pygmies appears to have been limited. However, female geneflow between both groups has been measured with mitochondrial DNA and the data results from this marker are rather interesting. According to the available mitochondrial DNA data (Quintana-Murci et al. 2008), female geneflow from the Bantus to the Pygmies has not occurred. However, significant female geneflow between Pygmies and Bantu farmers has occurred. This conforms to the anthropological record, as presented by Bahuchet (2012), whereby one "commodity" of Bantu and Pygmy trade was Pygmy brides. Perhaps the concept of hypergamy helps to explain language shift among the Pygmies. Hypergamy describes situations where women marry men of higher socio-economic ranking. For the Bantus and Pygmies this explains why Pygmy women married Bantu men, whereas marriage between Bantu women and Pygmy men had not occurred. Taking this a step further, the Pygmies could have shifted to the language of their closest Bantu neighbors as farmer languages were considered more prestigious. This, in turn, invites further research that identifies factors or aspects of farming cultures that create prestige among hunter-gatherers.

Prestige motivated language shift among the Pygmies merely represents a working hypothesis, which in turn, is posited to encourage researchers to further explore agriculture as a vehicle for language shift, a complex topic found not only in Africa, but elsewhere in the world. At this point some important caveats are in order. First, the amount of available genetic data, especially for Pygmies, are very limited. Additionally, the E1b-M180 mutation has also emerged as the genetic signature of the Bantu expansion (see Chapter 5). However, the frequency of this mutation among the Pygmy groups remains a mystery. Berniell-Lee et al (2009) report 20 percent based on small sample sizes from three populations. More data are clearly needed and more E1b-M180 data for the Pygmies may reveal greater male Bantu to Pygmy geneflow than what is suggested by the B2b-M150 mutation. This, in turn may suggest more intense contact between Bantus and Pygmies, an additional factor, in addition to prestige, that may have led to language shift.

Section 4. "Click Languages."

Click consonants are an exceedingly rare group of phonemes. Nevertheless, they stand as very productive speech sounds among the so-called Khoisan languages of southwestern Africa. Sandawe and Hadza, two of the isolate languages of Africa, also utilize click consonants. This raises an interesting question, whether Khoisan, Sandawe, and Hadza share a common linguistic history given the fact that clicks are a rare phoneme. Alternatively, since the Hadza and Sandawe live in Tanzania and are separated by over two thousand kilometers from the Khoisan, clicks may have evolved independently in several African languages.

The B2b-M112 mutation has evolved into an informative haplogroup for exploring the history of African click languages because it is found both in Khoisan populations as well as among the Hadza and Sandawe (see **Supplementary Table 3.2**). Knight et al. (2003) initially focused on B2b-M112 variation as found among the Hadza and mixed Khoisan samples. Based on their analysis of the B2b-M112 mutation and short tandem repeat (STR) variation, the study suggests that both populations diverged about 120 thousand years. However, this estimate has a huge margin of error of plus or minus 40 thousand years. Tishkoff et al (2007) broadened the examination of African click languages by including the Sandawe along with a mixed Khoisan sample and a sample from the Hadza. Based on analysis of the B2b-M112 mutation and short tandem repeat variation, they estimated that Khoisan, Sandawe, and Hadza diverged from an ancestral population about 35 thousand years ago. With a margin of error of plus or minus four thousand years, their estimate is more accurate than the one provided by Knight et al. (2003).

The significance of the studies by Knight et al. (2003) and Tishkoff et al. (2007) is that click consonants have significant time depth. These consonants could be a linguistic relic of the Middle Stone Age in Africa. Both studies reasoned that since click consonants are an especially rare phoneme, this speech sound had not evolved independently among the Hadza, Sandawe and Khoisan. Rather, click consonants stand as a linguist artifact of an ancient ancestral population from which all three populations descended. Both studies further suggest that click sounds initially evolved to aid hunters in their pursuit of game and later evolved into the earliest contrastive consonants.

Güldemann and Stoneking published a report in 2008 that questions the findings presented by Tishkoff et al. (2007). They assert that the researchers made conclusions based on insufficient information and other factors may account for the click consonants found in African languages, such as independent innovation or language contact. Support for the language contact argument certainly comes from Xhosa, a Niger-Congo language of South Africa that has click consonants in its phonemic inventory, which is unusual because these consonants are generally absent within this family of languages. An interesting study from 2016 (Rocha and Fehn) observed a high frequency of indigenous Khoisan mtDNA haplogroups among the Xhosa. This indicates that a substantial number of Khoisan women became members of Xhosa communities through marriage with Xhosa men. Taking this a step further, the mitochondrial DNA data support language contact as the source of click consonants in Xhosa language.

Geographical distance between Tanzania and South Africa undermines, however, language contact as an explanation for the presence of clicks in the Sandawe, Hadza and Khoisan. Moreover, the suggestion that clicks evolved independently among the Khoisan, Sandawe and Hadza also seems rather implausible. Contrastive click consonants are extremely rare and are only found in African languages. Some may cite Damin, a special register language found in Australia, as evidence of the potential of clicks to evolve independently. However, Damin is not classified as a language by *Ethnologue*, and even it was, clicks are still extremely rare. Thus, given the choice of a common ancestral language, language contact, or independent innovation, the most likely scenario is that Sandawe, Hadza and Khoisan all share a common ancestral language that has roots extending deep into the prehistory of language.

Section 5. Conclusions for Haplogroup B-M60.

The B2a-M150 and B2b-M112 mutations represent most of the published data for haplogroup BM60. These mutations are found in sub-Saharan Africa. Linguistically, the mutations stand as useful markers for deciphering the prehistory of the Khoisan macro-language family, the Niger-Congo language family, and the Hadza and Sandawe language isolates. Culturally, the B2a-M150 mutation stands a genetic relic of Bantu farmers their expansion from central western Africa roughly 5,000 years ago. B2b-M112, on the other hand, stands as the genetic signature of African hunter-gatherers. For linguists, the B2a-M150 and B2b-M112 markers present an opportunity to explore language contact theory and more specifically, agriculture as a vehicle for language shift. Finally, the B2b-M112 marker provides an opportunity to explore the history of the so-called "click" languages. Are click consonants the oldest phoneme?

Chapter 4: Haplogroup D-M174.

Section 1. The Evolutionary History of the D-M174 Mutation.

1.1. Overview.

The reader is invited to review **Supplementary Figure 1.1** from the first chapter. Both haplogroup E-M96 and haplogroup D-M174 diverge from the DE-M145 paragroup. Poznik et al. (2016) suggests that this occurred roughly 67 thousand years ago. Interestingly, the DE-M145 mutation was one of the first Y-chromosome polymorphisms that were discovered (Hammer 1994). One of the distinguishing characteristics of this mutation and its downstream variants, haplogroups D-M174 and E-M96, is the presence of a unique Alu insertion polymorphism. This explains why the literature sometimes describes DE-M145, D-M174 and E-M96 as positive for the Y Alu Polymorphism (or YAP+).

From **Supplementary Figure 1.1**, the reader will also observe that paragroup DE-M145 evolved from paragroup DR-M168, the ancestral mutation of all the "out-of-Africa" haplogroups. Besides haplogroups D-M174 and E-M96, the out-of-Africa haplogroups are C1-F3393, C2-M217, G-M201, H-M2713, I-M170, J-M304, L-M20, T-M184, M-P256, S-B254, N-M231, O-M175, Q-M242, and R-M207. The term "out-of-Africa" implies, of course, that these haplogroups are distinct from haplogroups A and B-M60, which evolved and remained in Africa. As such, haplogroups A and B-M60 convey the story of human evolution in Africa (see Chapters 2 and 3), whereas the out-of-Africa haplogroups stand as the genetic relics of the human colonization of Eurasia, Australia, the Americas, and Oceania that began roughly 60 thousand years ago.

1.2. Overview of Time Standards.

A discussion of the evolutionary history of the D-M174 mutation requires a discussion of the out-of-Africa exodus that occurred roughly 130 thousand years ago. The reader is now invited to examine Table 4.1 (below) which provides an overview of the time standards used to discuss the evolution of Y-chromosome diversity outside of Africa. The term "Marine Isotope Stage" is borrowed from the earth scientists and helps to carry a discussion of climate data. For example, improved climatic conditions at the beginning of Marine Isotope Stage 5, roughly 130 thousand years ago, facilitated the out-of-Africa exodus. Turning now to the term Holocene, this term comes from the field of geology and is useful for presenting useful data related to end of the Last Ice Age, roughly 12 thousand years ago. Finally, terms such as Paleolithic, Mesolithic, and Neolithic come from the field of archeology and are useful for data related to technological achievement. For example, agriculture appeared during the Neolithic and many of the major language expanded because this culture adaptation vastly improved reproductive success.

1.3. Overview of Paleoclimatological Concepts.

The following discussion of the Last Ice Age and the current Holocene epoch provides necessary background information that facilitates a discussion of contemporary Y-chromosome diversity. For example, the Native American cultural tradition is linked to the demise of large herbivores that occurred at the beginning of the Holocene.

Table 4.1. Overview of Time Standards that Carry a Discussion of the Prehistoric Data.

Date	**Earth Science**	**Geology**	**Archaeology**	**Comments**
130,000 years ago.	Marine Isotope Stage 5 begins.	Pleistocene continues from 2.6 million years ago.	Paleolithic continues from 2.6 million years ago.	The start of Marine Isotope Stage 5 coincides with the start of the Last Ice Age. The human tribe migrated out of Africa into the Levant.
71,000 years ago.	Marine Isotope Stage 4 begins.	Pleistocene continues.	Paleolithic continues.	The start of Marine Isotope Stage 4 coincides with the Toba eruption in Indonesia. A genetic bottleneck occurred among human populations.
57,000 years ago.	Marine Isotope Stage 3 begins.	Pleistocene continues.	Paleolithic continues.	Favorable weather conditions drive human expansions across Eurasia.
29,000 years ago.	Marine Isotope Stage 2 begins.	Pleistocene continues.	Paleolithic continues.	The start of Marine Isotope Stage 2 coincides with the Last Glacial Maximum. Late Pleistocene climate change drives human expansion into the Americas.
12,000 years ago.	Marine Isotope Stage 1 begins.	Holocene begins.	Mesolithic starts. Neolithic starts with the adoption of agriculture.	The start of Marine Isotope Stage 1, the Holocene, and the Mesolithic coincide with the end of the Last Ice Age. The Mesolithic is the time between the Paleolithic and Neolithic. The Neolithic marks the adoption of agriculture, the start of which various from region of the world to the next.

The term "ice age" denotes a period of glacial ice expansion across Northern Eurasia and lower sea levels worldwide because water is trapped within the ice. Glacial periods or ice ages are caused by variations of the earth's tilt and wobble in relation to its axis as well as temporary increases in distance between the sun and the earth. Thus, tilt, wobble and orbit are stable during so-called "interglacial periods" such as the current Holocene epoch that began 12 thousand years ago. Conversely, they are more unstable during glacial periods, including the Last Ice Age, which lasted between 130 and 12 thousand years ago. Taking this a step further, variations in the earth's tilt, wobble and orbit not only produced sudden fluctuations in the advance and retreat of glacial ice during the Last Ice Age, but also periods of precipitation or drought that appeared and disappeared within various regions of the world. This stands in sharp contrast to relatively stable climatic conditions in the current Holocene epoch because the earth's tilt, wobble and orbit are now stable. Consequently, the glaciers have retreated in the last 12 thousand years. As the glacial ice melted, the sea levels rose.

1.4. The Out-of-Africa Exodus.

Having now discussed prehistoric time standards as well as glacial and interglacial periods, a discussion of the "out-of-Africa" expansion now follows. Climatological, genetic, and archeological perspectives time a *successful* out-of-Africa expansion to the beginning of Marine Isotope Stage 5, roughly 130 thousand years ago. The use of *successful* helps to address archaeological evidence of *unsuccessful* out-of-Africa expansions that may have occurred earlier. Hershkovitz et al. (2018), for example, report the discovery of a human maxilla in Israel that is dated between 177 and 194 thousand years ago. Another study (Harvati et al. 2019) date a modern human cranium found in Greece to around 210 thousand years ago. These early expansions were *unsuccessful* because according to the available evidence, these individuals failed to contribute to contemporary genetic diversity among modern humans.

Turning now to the climate data, the beginning of Marine Isotope Stage 5 coincides with the beginning of the Last Ice Age. This signals the beginning of unstable weather conditions that potentially motivated the out-of-Africa migration. Data from a 2011 study (Frumkin et al. (2011) reports that at the beginning of Marine Isotope Stage 5 (roughly 130 thousand years ago), the Levant experienced greater precipitation. This period of greater precipitation extended to the beginning of the Holocene, roughly 12 thousand years ago. This not only explains why *Homo sapiens* migrated out of Africa, but also why they thrived and survived in this region during the last Ice Age.

Blome et al. in their 2012 paper provide additional evidence for an out of Africa migration during Marine Isotope Stage 5. Their study is a synthesis of a tremendous amount of paleoclimatological and archaeological data from all of Africa. The study divides the continent into four different regions for comparison purposes: Southern Africa, Tropical Africa, East Africa, and North Africa. Figure 15 of their study illustrates the density of archeological sites within these four regions from 150 to 30 thousand years ago, as well as periods of arid and humid conditions within this timeframe. This figure suggests that the beginning of Marine Isotope Stage 5 triggered dramatic climate change in Africa. Northern Africa experience greater precipitation whereas other regions experienced drought. This would have placed an extraordinary amount of pressure on hunter-gatherer populations in sub-Saharan Africa to migrate northwards in search of food.

A recent paper (Beyer et al. 2021) further elucidates the out-of-Africa expansion. The study reports computer simulations of ecological and anthropological data. One interesting observation is that the timing of a successful out-of-Africa expansion must be linked to sufficient precipitation. Hunter-gatherers require a minimum of 90 millimeters of rain per year to survive. Based on the climatological data, the study favors an out-of-Africa expansion via the Sinai Peninsula towards the beginning of Marine Isotope Stage 5. The study also reports two expansion windows during this period when the climatic condition would have optimal for the out-of-Africa migration, one about 130 thousand years ago, and another about 110 thousand years ago.

Fossil evidence also supports an out-of-Africa migration into the Levant during Marine Isotope Stage 5. Beginnings in the 1930s archaeologists have discovered several remains from Neanderthals and early modern humans from the Qafzeh and Skhul caves near the Sea of Galilee in Israel. According to Oppenheimer (2012), these remains are between 90 and 120 thousand years old. It should be emphasized that some researchers (e.g., Mellars 2006; Oppenheimer 2012) regard the Qafzeh and Skhul remains as evidence of an *unsuccessful* out-of-Africa migration during Marine Isotope Stage 5. However, recent paleoclimatological data presented above (Beyer et al. 2021; Blome et al. 2012; and Frumkin et al. 2011 offer especially persuasive arguments that redefines Qafzeh and Skhul remains as evidence of a successful out-of-Africa exit at the beginning of Marine Isotope Stage 5, which coincides with the onset of the last glacial period and greater precipitation in the Levant.

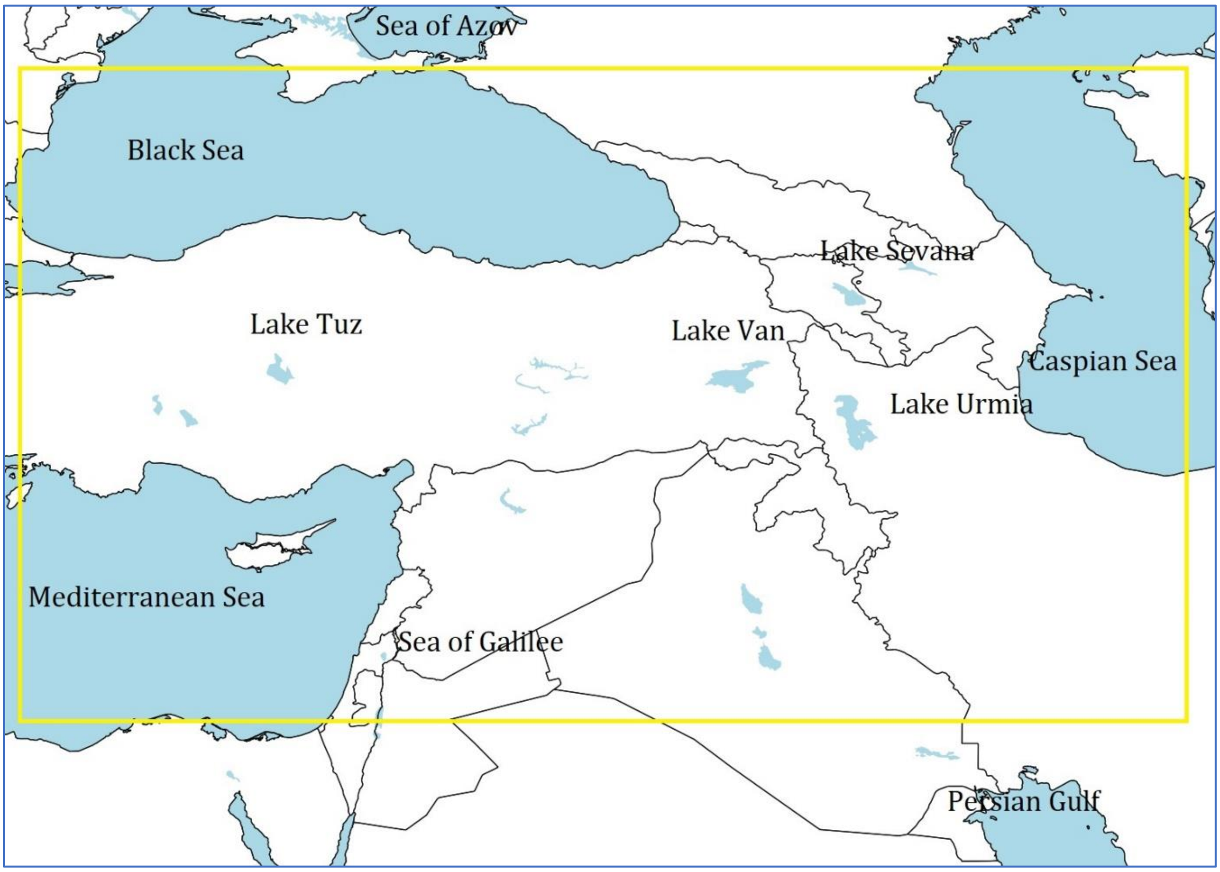
Turning now to the Y-chromosome evidence, as previously noted the DR-M168 paragroup is the ancestral mutation for all the out-of-Africa Y-chromosome haplogroups. Dating estimates from Poznik et al. (2016) suggest that the DR-M168 mutation evolved about 100 thousand years ago. This estimate correlates well with an out-of-Africa migration towards the beginning of Marine Isotope Stage 5, and with that, the beginning of Y-chromosome diversification in the Levant.

1.5. Hiatus in the Levant during Marine Isotope Stage 4.

At this point it is necessary to discuss geographical terminology. The reader is asked to review Figure 4.1 (below). For the purposes of the discussion in this chapter, the terms "Levant" and "Middle East" are synonymous. This area, which is surrounded by a yellow border in the figure below, begins at the Sea of Galilee and extends northwards to the southern coastline of the Black Sea and the Southern Caucasus region between the Black and Caspian Seas. The use of Middle East or Levant to describe this area is somewhat nonstandard but is necessary to deliver an efficient presentation of the data. Otherwise, the cumbersome term "putative homeland of non-African Y-chromosome diversification" must be used.

The reader is invited to follow the evolution of Y-chromosome diversity, from Adam to haplogroups D-M174 and E-M96, as diagramed in **Supplementary Figure 1.1** from the first chapter. Dating estimates from Poznik et al. (2016) suggest that DE-M145 evolved from DR-M168 about 70 thousand years ago. Later, about 67 thousand years ago, haplogroups D-M174 and E-M96 evolved from DE-M145. The evolution of DE-M145, D-M174 and E-M96 probably occurred in the Levant based on a synthesis of paleoclimatological, archeological, and genetic data. As noted previously, Frumkin et al. (2011) indicate that climatic conditions in this region were conducive for population growth during Marine Isotope Stage 4. Turning now to the fossil record, a partial modern human skull dated to at least 55 thousand years ago was found at the Manot Cave in Israel. This skull has unique Neanderthal features that are absent from the remains at Qafzeh and Skhul in Israel which as previously noted, came from people who died between 90 and 120 thousand years (see Hershkovitz et al. 2015 for more details).

Figure 4.1. The Putative Homeland of Non-African Y-Chromosome Diversification (the area enclosed by the yellow rectangle).



To understand the significance of the Neanderthal features found in the remains from Manot Cave, it should be noted that Green et al. (2010) report that between one and four percent of the human genome consists of DNA inherited from the Neanderthals, archaic hominids who became extinct about 30,000 years ago. Apparently, some mating (admixture) occurred between them and a small number of early human modern humans (e.g., Currat et al. 2011). Additionally, Green et al. (2010) report that Neanderthal DNA is found only found in non-Africans. As such, they suggest that admixture between Neanderthals and Homo sapiens occurred in the Levant, a conclusion supported by the fossil record and the known range of Neanderthals. Finally, Green et al (2010) suggest that human and Neanderthal admixture occurred before separation of the ancestral population of Eurasians and the Aboriginal Australians. This stems from the observation that all non-African populations have "statistically indistinguishable" amounts of Neanderthal DNA (Reich et al. 2011).

Evidence of Neandertal and human admixture in the Levant during Marine Isotope Stage 4 undermines a recent study (Hallast et al. 2021) that identifies East Asia as the origin of non-African Y-chromosome diversity which includes haplogroup D-M174. As a matter of fairness, it should be noted that their position is potentially supported by the archeological record. Liu et al., in their 2015 paper, report of the discovery of forty-seven human teeth that they confidently date between 80 and 120 thousand years ago. These remains were found at the Fuyan Cave in southern China. Westaway et al. (2017) report fossil human teeth found at the Lida Ajer cave in Sumatra that are dated between 73 and 63 thousand years ago. Nevertheless, the position taken by Hallast et al. (2021) is problematic because East Asia is outside the known range of Neanderthals. For this reason, contemporary non-African Y-chromosome mutations could not have evolved in this region of the world. Rather, the remains from Fuyan Cave and Sumatra belonged to individuals who failed to contribute to contemporary human genetic diversity. Perhaps this can be explained by populations bottlenecks and extinctions that occurred after the catastrophic Toba volcano explosion in Indonesia about 75 thousand years ago (for more details, see Rampino et al. 2000; Wei and Li 2017).

Interestingly, the Y-chromosome data in Table 4.2 (below) suggest human populations in the Levant experienced a bottleneck because of the Toba eruption. Bottlenecks describe a re-shaping of genetic variation that occurs after a near-extinction event. The large gap found between the evolution of the BR-M42 and DR-M168 mutations, about 50 thousand years, as well as the gap between DR-M168 and the evolution of non-African Y-chromosome diversity that followed, about 30 thousand years, suggests partial obliteration of the genetic trail that unites Africans and Eurasians. Furthermore, the rapid diversification of non-African mutations that occurred roughly 67 thousand years suggests the following: populations in the Levant recovered from the near-extinction event within a period of about 10 thousand years.

1.6. Expansion of *Homo sapiens* across Eurasia via a Southern Route.

Pope and Terrell in their 2008 paper provide environmental context for a southern coastal migration to from the Levant to East Asia and beyond during Marine Isotope Stage 3. According to the paper the Indian sub-continent experienced cold and dry conditions during Marine Isotope Stage 4, which also produced wild fluctuations in sea level. Such conditions would not have provided sufficient marine resources to fuel a human migration along southern Asian coastline. However, warmer weather and the monsoon rains returned during Marine Isotope Stage 3 and the sea level became stable. According to Pope and Terrell (2008), improved climatic conditions eventually facilitated this expansion.

One question posed by the southern dispersal hypothesis is how the human tribe reached the Gulf of Oman to begin their coastal migration. Traditionally, researchers have assumed that the out-of-Africa migrations entered the Arabian Peninsula via the Red Sea and the narrow Gate of Tears that separates East Africa and Yemen. Then the migration followed the southern coast of the Arabian Peninsula and crossed over into South Asia at the Straits of Hormuz, another narrow crossing point (e.g., Mellars 2006; Oppenheimer 2012). One problem with this traditional model is that the fossil and archaeological record fails to support human occupation of the southern Arabian Peninsula during the Last Ice Age. (e.g., Bailey et al. 2007). Additionally, dispersal from the Arabian Peninsula model is inconsistent with the paleoclimatological data. During the Last Ice Age, the lack of precipitation would have hindered human occupation of the region (see Beyer et al. 2021; Parton et al. 2015). Perhaps a more parsimonious scenario is that modern humans were drawn towards the Black and Caspian Seas during Marine Isotope Stage 4. With the onset of warmer weather at the beginning of Marine Isotope Stage 3, groups of modern humans migrated alongside the Euphrates and Tigris Rivers to the Persian Gulf, and then along the coastline of South Asia to points beyond.

Table 4.2. Evolution of Non-African Y-Chromosome Diversity. Mutations highlighted in green are haplogroups. Sources: With the exception of haplogroups S-B254 and M-P256, time estimates are from Poznik et al. (2016: Supplementary Table 10). Estimates for haplogroups S-B254 and M-P256 are from Karmin et al. (2015). Note: Estimates for the evolution of Y-chromosome diversity always represent a "ballpark figure." They help to decipher what is and what is not plausible. The region where a Y-chromosome mutation evolved is based on my analysis of the data.

Mutation	Average Pedigree Time from Poznik et al. 2016 (years)	Marine Isotope Stage (MIS)	Where?
BR-M42	147,550	MIS 6	Africa
DR-M168	98,150	MIS 5	Middle East
DE-M145	70,500	MIS 4	Middle East
CR-P143	70,500	MIS 4	Middle East
C-M130	70,050	MIS 4	Middle East
FR-M89	70,050	MIS 4	Middle East
E-M96	67,450	MIS 4	Middle East
D-M174	67,450	MIS 4	Middle East
G-M201	50,300	MIS 3	Middle East
H-M2713	50,100	MIS 3	Middle East
IJ-M429	49,250	MIS 3	Middle East
LR-M9	49,250	MIS 3	Middle East
C1-F3393	48,700	MIS 3	Middle East
C2-M217	48,700	MIS 3	Middle East
KR-M526	47,250	MIS 3	Middle East
LT-L298	47,250	MIS 3	Middle East
NO-M214	47,150	MIS 3	Central Asia
P-M45	47,150	MIS 3	Northern Eurasia
S-B254	46,750	MIS 3	Island Southeast Asia
M-P256	46,750	MIS 3	Island Southeast Asia
T-M184	44,650	MIS 3	Middle East
L-M20	44,650	MIS 3	Middle East
J-M304	44,150	MIS 3	Middle East
I-M170	44,150	MIS 3	Europe
N-M231	41,500	MIS 3	East Asia
O-M175	41,500	MIS 3	East Asia
Q-M242	32,500	MIS 3	Northern Eurasia
R-M207	32,500	MIS 3	Northern Eurasia

As noted previously, the Y-chromosome data suggest that by around 65 thousand years ago the hunter-gatherers in the Levant recovered from the Toba catastrophe. From an anthropological perspective, it seems plausible that some of these hunter-gatherers may have gravitated towards the contemporary Black Sea. According to Yanchilina et al. (2019), the Black Sea was a large freshwater lake at the beginning of Marine Isotope Stage 3. Taking this a step further, this body of water potentially supported the food resources needed by hunter-gatherers. Moreover, these resources would have fueled rapid population growth that occurred after the Toba eruption, an observation that is supported by the Y-chromosome data (see Table 4.2 above).

Core sample data from taken from the Black Sea may explain why moderns migrated away from this region to East Asia and elsewhere during Marine Isotope Stage 3. According to Wegwerth et al. (2021), climate data suggest that about 60 thousand years ago, around the onset of Marine Isotope Stage 3, the climatic condition in northern Anatola underwent a major transformation. A warming cycle caused the ice glaciers across Eurasia to retreat. More rainfall and meltwater from the retreating glaciers caused the Black Sea to increase in size. The steppes of northern Anatolia became forests. Conceivably, this change in climate may have motivated some of them to expand out of northern Anatolia. Perhaps climate change caused rapid population growth, and perhaps population pressure forced hunter-gatherers to find food elsewhere. Alternatively, climate change may have resulted in diminished food resources and a need to expand into region with more abundant resources; or perhaps hunter-gatherers migrated away from the Black Sea to escape flooding.

Marine Isotope Stage 3 began 57 thousand years ago and ended about 29 thousand years ago. During this period, part of the human tribe remained in the Levant. Some, however, migrated eastwards, and some migrated westwards. Eastward expansions during this stage resulted in the human settlement of South Asia, East Asia, and Australia. Westward expansions include the human colonization of Europe and a back-migration to Africa. It should be noted that researcher have traditionally favored a single "southern dispersal" coastal route that defines the human expansion into East Asia and Australia (e.g., Mellars 2006; Stoneking and Delfin 2010; Oppenheimer 2012). However, recent genetic and archeological evidence suggest a second "northern route," as topic that will be discussed in greater detail in Chapter 6.

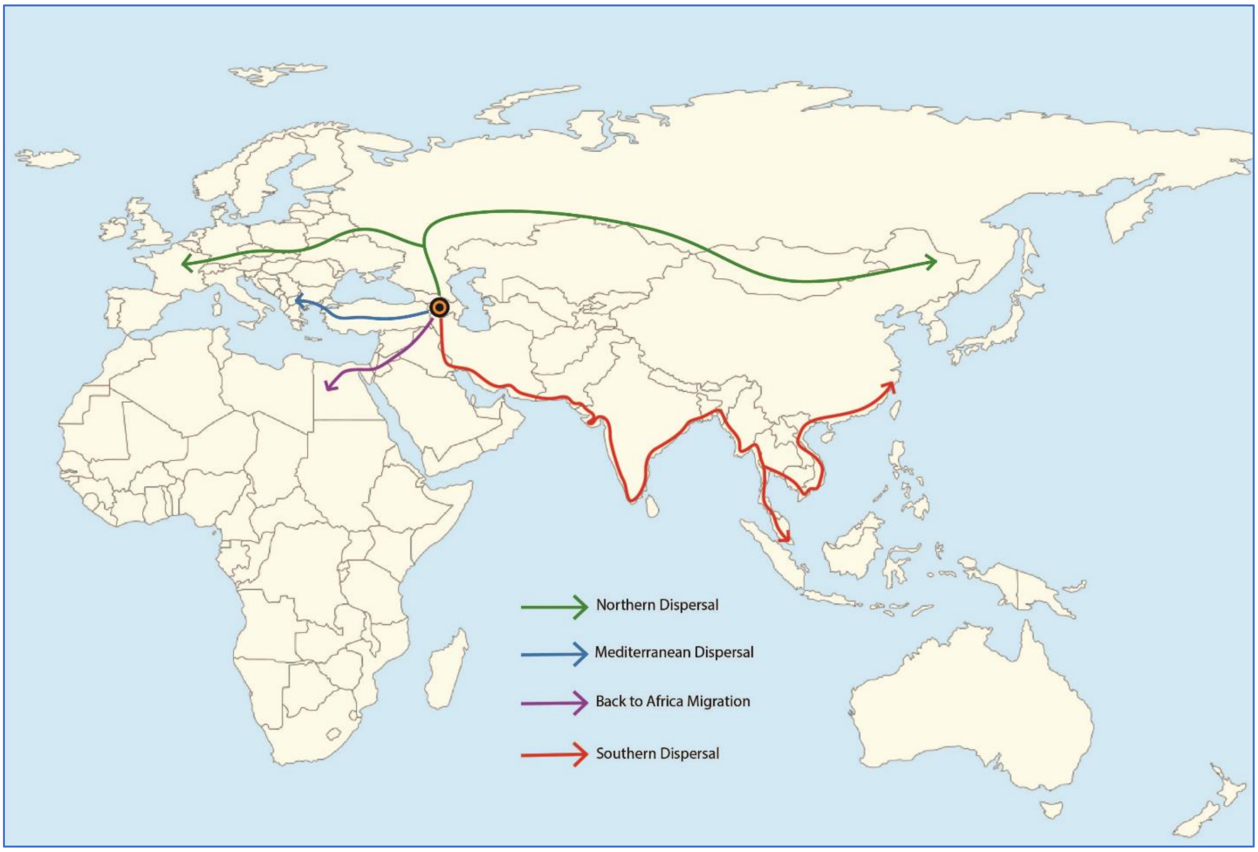
Turning now to the southern dispersal route, Pope and Terrell (2008) suggest that this migration initially expanded eastwards along the south Asian coastline. Important archeological support for this position comes from artifacts and human remains that were discovered at the Fa-Hien Lena Cave in Sri Lanka (Wedage et al. 2019). These data support the presence of *Homo sapiens* in South Asia to at least 45 thousand years ago. Pope and Terrell (2008) also suggest that the southern dispersal migration into in southeastern Asia encountered a geographic cul-de-sac at the Pacific Ocean. Archaeological support for this position comes from modern human remains found at the Ta Pa Ling Cave in northern Laos. According to Demeter et al. (2012) these remains come from an individual who died roughly 45 thousand years ago.

After reaching the geographic cul-de-sac in southeastern Asia, some migrated southwards into Sunda and Sahul, and then into Australia. Important archaeological support for this expansion comes from Lake Mungo in Australia and modern human remains that are dated to at least 46 thousand years ago (see Bowler et al. 2003). Genetic relics of the expansion into Sahul are downstream variants of haplogroups C1-F3393 (see Chapter 6), as well as haplogroups M-P256, and S-B254 (see Chapter 14). However, unlike haplogroups C1-F3393, M-P256, and S-B254, populations with the D-M174 mutation also expanded northwards from southeastern Asia along the Pacific coastline.

1.7. Expansion of D-M174 and *Homo sapiens* into East Asia.

The archeological record and ancient Y-chromosome data identify haplogroup D-M174 as a genetic relic of modern human expansions into East Asia via the southern route. The available ancient Y-chromosome data is summarized in **Supplementary Table 4.1**. Part of the data in the table was reported by McColl et al. (2018). The researchers sequenced haplogroup D-M174 from the remains of an individual who died in Malaysia roughly four thousand years. Culturally, the remains are associated with the Hoabinhian tradition. Higham (2013) provides a useful overview of this tradition. The term *Hoabinhian* represents a generic description of hunter-gatherers in southeastern Asia who thrived in this region prior to the beginning of the Neolithic. Moreover, archaeological evidence of the Hoabinhian tradition often appears at rock shelters found throughout this region. A study from 2016 (Ji et al.) further notes that the term evolved roughly one hundred years ago among French archeologists who initially discovered evidence of this tradition in the Hoa Binh province of northern Vietnam. According to the study, artifacts linked with this tradition include unique stone tools shaped with cobbles.

Figure 4.2. Human Migration from the Levant during Marine Isotope Stage 3.



In their 2016 study, Ji et al. report data gathered from the Xiadong archaeological site located Yunnan province of southern China. This site is near the Mekong River, where China, Myanmar, Thailand, and Laos converge on a map. Based on radio-carbon dating results and artifacts gathered from the site, the researchers claim that Hoabinhian hunter-gatherers were in southern China around 40 thousand years ago. Moreover, the study claims that the Xiadong archaeological site records the earliest presence of the Hoabinhian tradition in southeastern Asia.

Besides the Hoabinhian tradition, the ancient Y-chromosome data (see **Supplementary Table 4.1**) suggest that haplogroup D-M174 stands as a genetic relic of the hunter-gatherers who colonized the Japanese archipelago. As previously discussed, the southern dispersal encountered a geographic cu-de-sac roughly 45 thousand years ago in southeastern Asia. According to Pope and Terrell (2008), those that expanded northwards along the Pacific Ocean eventually reached the Korean Peninsula roughly 38 thousand years ago. In a 2012 study, Takashi provides radiocarbon dating results from archeological sites in Japan. He also details the distribution of stone axes and other tools. According to the researcher, these data suggest that hunter-gatherers expanded from Korea to the Japanese island of Kyushu roughly 35 thousand years ago, a remarkable feat as this required an ocean crossing with primitive watercraft. Shortly thereafter, these Paleolithic foragers expanded from Kyushu to Honshu, the largest of the Japanese islands.

Hudson (2009 and 2017a) provides a useful overview of the prehistoric Jomon culture of Japan. This hunter-gatherer cultural tradition evolved roughly 16 thousand years ago. This era in the Japanese prehistory includes the use of pottery, a greater dependence on marine resources for food, and more permanent settlement patterns, which is unusual among Paleolithic hunter-gatherers. Populations that forage are generally mobile. As shown by **Supplementary Table 4.1**, researchers have sequenced haplogroup D-M174 from Jomon remains found on the island of Kyushu and on Hokkaido. These remains link haplogroup D-M174 with the arrival of modern human in Japan, roughly 35 thousand years ago.

Section 2. Contemporary Distribution of Haplogroup D-M174.

2.1. Overview.

As shown by **Supplementary Table 4.2**, haplogroup D-M174 is found almost exclusively in East Asian populations. In this region, the overall frequency of D-M174 is around 11 percent (e.g., Zhong et al. 2011). Moreover, the distribution of D-M174 exhibits an unusual frequency pattern in that heavy frequencies are concentrated in Tibet and Japan, which represent the western and eastern limits of East Asia. Outside of East Asia, the D-M174 haplogroup represents attains a significant frequency among Andamanese Islanders.

The reader is directed to **Supplementary Figure 4.1**, which provides a phylogenetic diagram of the D-M174 haplogroup. Turning now to linguistically informative variants of D-M174, mutations downstream from D1a-F6251 help to decipher the prehistory of Tibeto-Burman languages (see, also, see **Supplementary Table 4.3**). D1b-M55 mutations help to decipher the prehistory of Japonic and the Ainu language isolate (see **Supplementary Table 4.4**). Finally, the D1c-Y34638 mutation helps to decipher the prehistory of the Andamanese languages (see **Supplementary Table 4.5**).

2.2. Haplogroup D-M174 and Japonic.

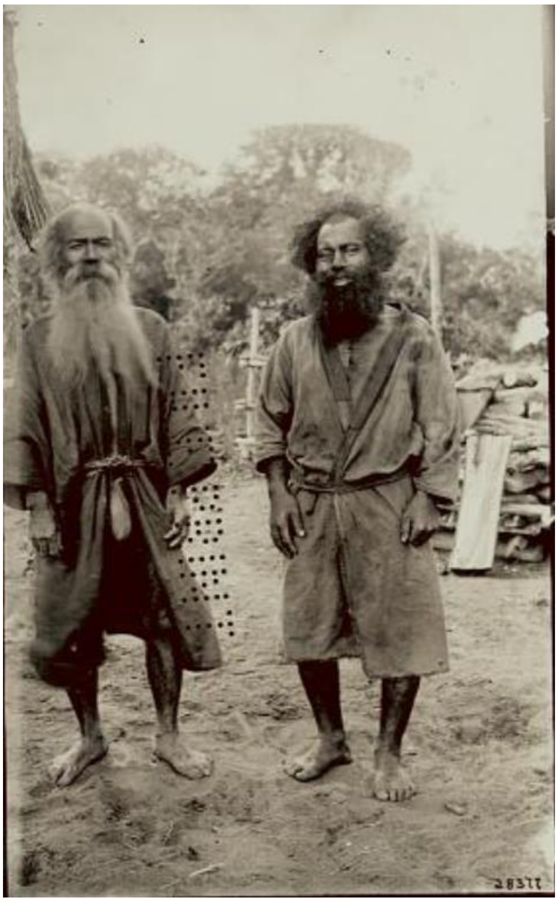
As mentioned earlier, Paleolithic hunter-gatherers arrived in the Japanese Islands roughly 35 thousand years ago. Among the geneticists the D-M174 haplogroup has been identified as Paleolithic component among contemporary Japanese (e.g., Hammer et al. 2006). Moreover, this Paleolithic contribution to the contemporary Japanese pool is significant. Sato et al. (2014) conducted a large-scale study of over 2,000 Japanese males and found that 32 percent have the D-M174 mutation. Within the D-M174 main haplogroup, almost all the genetic variation among contemporary Japanese consists of the downstream D1b-M55 marker. This marker appears to have evolved in Japan based on absence of the mutation elsewhere in the world and the identification of the mutation in samples from archeological sites (see **Supplementary Tables 4.1** and **4.4**).

Besides D1b-M55, the C1a1-M8 mutation also stands as a Paleolithic component of the contemporary Japanese gene pool. Accordingly, this discussion continues in Chapter 6.

2.3. Haplogroup D-M174 and the Ainu Language Isolate.

The Ainu are an indigenous hunter-gatherer culture from the northern Japanese Island of Hokkaido, the Russian island of Sakhalin, and the Kurile archipelago, which is now under Russian administration. They speak Ainu which is classified as a language isolate. According to De Graaf and Shiraishi (2013), the Ainu appeared in the historical record roughly 300 years ago when Dutch explorers conducted a mapping survey of northern Japan. As the result of subsequent Japanese expansions into Hokkaido, and twentieth century wars between the Japan and Russia, the number of Ainu have dwindled to about 25 thousand people on Hokkaido. Additionally, about 20 years ago the Ainu culture finally attained legal protection in Japan. As such, their language faces an uncertain future due to widespread use of Japanese on Hokkaido.

Figure 4.3. Two Ainu Men, Tokoro Village, Hokkaido Island, Japan, between 1885 and 1895. Source: the Smithsonian Institution.



Y-chromosome data for the Ainu was reported by Tajima et al. (2004) and Hammer et al. (2006) and unfortunately, the Y-chromosome history of this population must be drawn from just 20 samples. Most of the Ainu have the D1b-M55 mutation, which represents a Paleolithic component of the contemporary Japanese gene pool (**see Supplementary Table 4.4**). Unlike the Japanese however, the available data also suggest a significant presence of the C2-M217 mutation among the Ainu. Accordingly, the discussion of the Ainu continues in Chapter 7.

2.4. Haplogroup D-M174 and the Andaman Islands.

The term "Negrito" surfaces in published studies to describe several small, isolated populations of contemporary Asia. Their appearance apparently resembles that of the African Pygmies; hence, the Spanish word for "black" with the diminutive affix. The Negrito populations of Asia include the Jarawa and Onge of the Andaman Islands, the Semang of Malaysia, the Maniq of Thailand, and the Aeta of the Philippines. Because of their unique appearance (or phenotype), some researchers have taken an interest in the Negritos to determine if they are a relic population from the out-of-Africa expansion during Marine Isotope Stage 3 (see Higham 2013 for additional details).

The Andaman Islands are found in the Bay of Bengal, which is part of the Indian Ocean. This location represents part of the southern dispersal route followed by the eastward out-of-Africa expansion. The Negritos of this archipelago, or the Andamanese, consist of four tribes: the Great Andamanese, Onge, Jarawa, and Sentinelese. Venkateswar (1999) provides a useful anthropological discussion of these cultures. She suggests that the arrival of modern humans in the Andaman Island may have occurred during the last Ice Age when sea levels were lower. However, the Andamanese only appeared in the historical record during the mid-19th century when the British arrived in the Andaman archipelago to construct a prison. Since then, disease and loss of territory have reduced the population of Andamanese from 5,000 in 1850 to around 500 people in 1999.

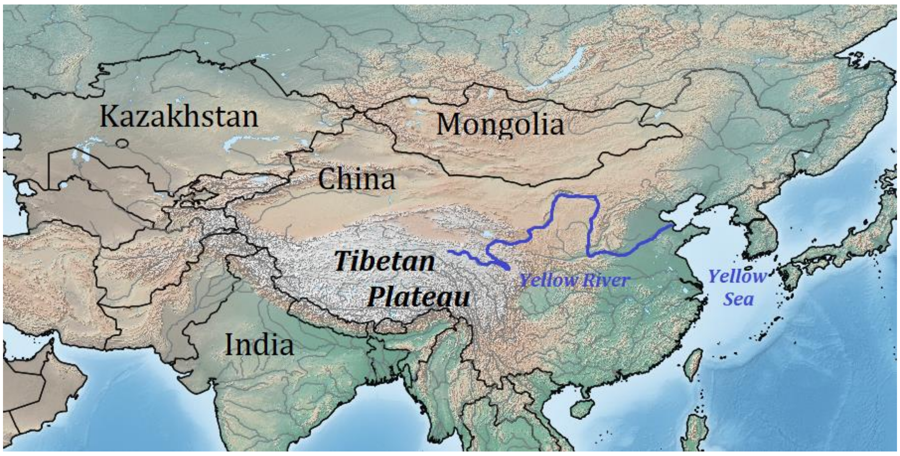
Figure 4.4. Andaman Islanders about 1911. Source: Smithsonian Institution.

A study from 2003 (Thangaraj et al.) tested 23 Onge samples and four Jarawa samples and found that all the samples contained unspecified variant of haplogroup D-M174. The International Society for Genetic Genealogy (2018 and 2019/2020) later defined this unspecified variant as the D1c-Y34638 mutation (see, also, **Supplementary Table 4.5**). In 2017, Mondal et al. reported data for the Onge and Jarawa based on whole genome sequencing. According to the study, contemporary Japanese and Andamanese Islanders separated from a common ancestral population around 53 thousand years ago. For the anthropologists, these data help to confirm that the human tribe followed the southern dispersal route when they migrated to East Asia about 50 thousand years ago. For the linguists, of course, the D1c-Y34638 mutation stands as a useful marker for exploring the prehistory of Andamanese languages.

2.5. Haplogroup D-M174 and Tibeto-Burman Languages.

The Tibetan Plateau encompasses most of the historical region of Tibet. Here the average altitude is around 4,000 meters above sea level. According to a 2016 report (Zhang et al.) the inhabitants of this region possess a unique evolutionary adaptation that enables them to survive at such an extreme altitude. The same study further suggests, based on archaeological remains, that nomadic hunter-gatherers first inhabited the plateau around 30 thousand years ago on a seasonal basis. The transition to permanent settlement of the region began about 7,000 years ago when farmers began to cultivate millet in the Middle Yellow River region of China. A thousand years later, millet cultivation expanded westwards to the northeastern rim of Tibetan Plateau. However, according to Zhang et al. (2016) the interior of the region was left to the hunter-gatherers until around 3,600 years ago. At this point farmers began to cultivate barley on the Tibetan Plateau, a crop that is more resistant to cold and dry climate of the region. Additionally, the success of agriculture in this region can be explained by the domestication of the Yak, a type of bovine. According to the genetic evidence, this occurred by around 7,000 years ago (Qiu et al. 2015). Among the Tibetans, this animal became an important source of meat, clothing, fuel, and transport.

Figure 4.5. The Tibetan Plateau and Yellow River.



Qi et al. in their 2013 study report data gathered from over 2,000 Tibetan males. According to the study around 54 percent of the sequences belong to haplogroup D-M174 and 33 percent belong to O-M175. Haplogroup O-M175 represents a Neolithic component among the Tibetans, a genetic relic of the westward expansion of agriculture into the region from China (see Chapter 16). Haplogroup D-M174, on the other hand, reflects a much earlier hunter-gatherer component, a relic of the human colonization of East Asia. As noted earlier, the D1b-M55 mutation is a unique Japanese variant of haplogroup D-M174. The D1c-Y34638 mutation, on the other hand, represents the genetic signature of Andaman Islanders. Similarly, Tibetans also have their own haplogroup D-M174 variant, the D1a-P99 mutation. According to Qi et al (2013), the D1a-P99 mutation evolved about 19 thousand years ago towards the end of the Last Ice Age. Among Tibetans the most common variant of D1a-P99 is the D1a-P47 mutation. Data from Qi et al. (2013) suggest that the D1a-P47 mutation evolved around 10 thousand years ago which correlates well with the evolution of agriculture in East Asia.

2.6. D1a-F6251, Hmong-Mien, and Tai-Kadai.

As shown by **Supplementary Figure 4.1**, the D1a-F6251, D1b-M55, and D1c-Y34638 mutations form linguistically informative lineages within the D-M174 haplogroup. The most resolved mutations downstream from D1a-F6251 are reported for Tibeto-Burman speakers, the D1a-P99 and D1a-P47 mutations. D1a-F6251 data for Hmong-Mien and Tai-Kadai speakers, on the other hand, remain poorly resolved. Although the D1a-F6251 mutation is frequently observed in these populations, the frequency numbers are generally less than 10 percent and as such, researchers have little incentive to find informative downstream variants (see, also, **Supplementary Table 4.3**). Nevertheless, based on the available data, D1a-F6251 mutation is a Paleolithic component of East Asian populations that managed to survive the Neolithic and with that, a massive expansion of O-M175 mutations. According to researchers (Qi et al. 2013; Wang et al. 2013b), around 8,000 years ago Neolithic farmers in East Asia had predominately haplogroup O-M175 while hunter-gatherers were predominately haplogroup D-M174. A massive expansion of farmers during the Neolithic then "shoved" haplogroup D-M174 to the periphery of East Asia (see Chapter 16 for additional details).

Section 3. Purported African Origins of D-M174.

Shang et al. (2007) report discovery of the so-called Tianyuan Man, remains found inside a cave near Beijing. According to radiocarbon dating results, he died about 40 thousand years ago. Moreover, the study reports that the remains consist of 34 bone fragments, which represent a remarkable amount of fossil data from a Paleolithic individual. Dental morphology, for example, suggests that he was about fifty years old at the time of his death. Moreover significantly, Shang et al. (2007) take the position that his skeletal morphology fails to a rapid and direct colonization of East Asia by modern humans from Africa. Rather, the morphology suggests that modern human populations underwent a long period of genetic diversification after leaving Africa. Ancient autosomal data harvested from Tianyuan also supports this interpretation of the skeletal morphology (see Yang, Melinda et al. 2017). These data support the position taken earlier, that before prior to the successful human colonization of East Asia (roughly 50 thousand years ago), genetic diversification of modern humans had occurred in the Levant.

Evidence of a long period of genetic diversification in Southwest Asia during Marine Isotope Stage 4, as provided by Tianyuan, undermines a 2019 study (Haber et al.) that posits the evolution of haplogroup D-M174 in Africa. Their findings are based on the presence of an exceedingly rare D0 mutation that was detected in five men from Nigeria. This model is problematic because researchers would have to assume that the successful out-of-Africa migration occurred 50 to 59 thousand years ago. This, in turn, would posit that the expansion from Africa to East Asia was rapid, a conclusion that is clearly inconsistent with Tian Yuan's skeletal morphology. Additionally, the position taken by Haber et al. (2019) is problematic because it fails to account for the difference in Neanderthal DNA found in non-African versus African populations. A recent computer simulation study (Chen et al. 2020) reports that the small amount of Neanderthal DNA detected in African populations is best explained by back-migration to the continent. The Y-chromosome relic of this back-migration is haplogroup E-M96 (see Chapter 5).

Section 4. The Evolution of Language, Climate Change, and Natural Selection.

The survival of the animal and plant life of our plane tends to be linked to a very limited ecological niche that offers opportunities and imposes constraints. As such, the survival of flora and fauna is sensitive to climate change. For example, an alligator in the Florida everglades is ill-equipped to survive on the frozen arctic icepack. Similarly, a polar bear from the Arctic Circle is ill-equipped to survive in the everglades. This discussion of the general trend in ecology leads to an interesting observation. Unlike most other animals, or even the archaic species of our genus, such as *Homo erectus* or Neanderthals, *Homo sapiens* can occupy a comparatively broad ecological niche.

The climate driven migration out of Africa, roughly 130 thousand years ago, pushed us into new biomes. Our cognitive abilities forged cultural adaptations that exploited the new opportunities that they offered and overcame the limitations that they imposed. Climatological and archaeological support for this position is outlined in paper published in 2018 by Roberts and Stewart. The researchers detail successful human adaptation to climate extremes that include deserts, rainforests, high altitude, and the arctic tundra. For example, the Paleolithic mammoth hunters of Northern Eurasia perfected the hunting of large herbivores. As such, they successfully exploited the food resources of the tundra steppes. Furthermore, they utilized mammoth bones as fuels, which overcame a limitation of this region, cold climatic conditions.

Roberts and Stewart (2018) further explain that “ecological plasticity” among humans was facilitated by highly advanced collaborative problem-solving skills. According to the researchers, when human migrated out of Africa into new ecological environments, no single person could possess all of the requisite knowledge for adapting to climate change. However, the tribe could collectively develop an adaptive response and then transmit this knowledge to the next generation. Taking this a step further, language plays a huge role in the evolutionary success of *Homo sapiens*.

As previously mentioned in Chapter 2, recent archeological studies support the idea that *Homo sapiens* had the cognitive ability to form language when they migrated out of Africa. Building on this observation, the ability to formulate language also has an anatomical component. Unlike cognitive milestones, however, the archeological record cannot measure the evolution of speech anatomy because soft tissue decays rapidly after death.

DNA methylation is a biochemical process that can retard the full expression of a gene without changing the sequencing of the gene. Perhaps a useful analogy is a commercial vehicle retrofitted with a speed governor. The motor and transmission can propel the vehicle to 140 km/h, but the governor limits road speed to 90 km/h in order to save fuel. A recent study (Gokhman et al. 2020) utilized a novel approach to investigate the evolutionary history of human speech anatomy. The study builds upon research that identifies the SOX9, ACAN, COL2A1, NFIX and XYLT1 genes as those that control the anatomy of the face and larynx. Gokhman et al. (2020) were able to compare the methylation patterns on these genes from contemporary and ancient samples taken from modern humans, as well samples from a Neanderthal, a Denisovan (a species of ancient humans), and chimpanzees. The study found that *Homo sapiens* have the most extensive changes related to DNA methylation. Among modern humans, these changes have caused the face to retract and the larynx to lower to such a degree that we can create the wide range of contrastive phonemes that are attested in the languages of the world.

The data from Gokhman et al. (2020) supports the following argument: language requires the anatomical infrastructure that evolution has only given to the modern humans. Their data, along with climatological, archeological, and other genetic perspectives, clearly suggest that when the human tribe migrated out of Africa, they were equipped with an advanced communication tool. This "tool" facilitated novel adaptations to climate change.

Section 5. Archaic Humans and the Y-Chromosome of *Homo Sapiens*.

Since the weight of genetic and archeological opinion currently favors admixture between humans and Neanderthals, this hypothesis will be utilized for the out-of-Africa model presented above in Section 1 of this chapter. However, it should be noted that William Amos, a researcher at Cambridge University in England, has recently questioned the reliability of this hypothesis. The admixture hypothesis is based on complex statistical models that examines African and non-African mutation rates (see Sankararaman 2020 for an overview). According to Amos (2020) and Amos (2021), researchers have been using the wrong statistical model to generate these data.

Admixture between Neanderthals and modern humans is also not supported by uniparental markers. Krings et al. (1997) reports, based on a comparison of modern human and Neanderthal mtDNA, that both populations underwent reproductive isolation. A similar conclusion was reached by Mendez et al. (2016) by comparing the Y-chromosome of Neanderthals with that of modern humans. One explanation for the mitochondrial and Y-chromosome DNA results is that Neanderthal uniparental markers were eliminated from the human genome as the result of genetic drift. Neanderthal autosomal DNA, on the other hand, may have boosted the immune system among modern humans and consequently was not eliminated from the genome by recombination (see Abi-Rached et al. 2011). A similar argument was made by Sankararaman et al. (2014) with the suggestion that Neanderthal autosomal DNA allows modern human to synthesize vitamin D at higher latitudes.

Section 6. Conclusions for Haplogroup D.

A discussion of non-African Y-chromosome data, which includes haplogroup D-M174, requires a *working* out-of-Africa model. The term "working" concedes that the model is subject change. For example, a future model might deem the human/Neanderthal admixture data as unreliable. This working model utilizes time standards from earth sciences, geology, and archeology to carry a discussion of the evolution of non-African Y-chromosome diversity. Based on the currently available data, around 130 thousand years ago the human tribe expanded into the Levant. Haplogroup D-M174 probably evolved in this region about 67 thousand years ago. About 50 thousand years ago, hunter-gatherers carried this mutation into South and East Asia. These hunter-gatherers include the prehistoric Hoabinhian tradition of Southeast Asia and the Jomon culture of Japan. Today, the contemporary distribution of haplogroup D-M174 mutations provides useful markers for deciphering the prehistory of Andamanese, Japonic, and Tibeto-Burman languages as well as the Ainu language isolate. These languages have deep roots that extend into the Paleolithic.

The discussion of the out-of-Africa expansion in this chapter provides an opportunity to examine the evolution of language as a behavioral adaptation among *Homo sapiens*. This adaptation probably occurred in Africa. As we expanded across Eurasia and beyond, language gave human tribe "ecological plasticity." Our highly advanced collaborative problem-solving skills then enabled us to survive in a wide range of environmental conditions. Finally, the out-of-Africa discussion provides an opportunity to address admixture between humans and Neanderthals. Is this really a case of smoke and mirrors as suggested by William Amos on social media?

Chapter 5: Haplogroup E-M96.

Section 1. Contemporary Distribution of Haplogroup E-M96.

Despite phylogenetic closeness, the geographic distribution of E-M96 and D-M174 are very much different. As explained in the previous chapter (Chapter 4), haplogroup D-M174 plays a rather modest role in representing the genetic diversity of East Asia. Haplogroup E-M96, on the other hand, represents a significant genetic marker for deciphering the evolutionary history of populations in Mediterranean Europe, southeastern Europe, the Middle East, and North Africa. Additionally, haplogroup E-M96 represents almost all the Y-chromosome genetic diversity in Sub-Saharan Africa, where over 92 percent of men have a variant of this haplogroup (Luis et al. 2004).

For linguists, E-M96 mutations help to decipher the prehistory of the Afro-Asiatic, Nilo-Saharan, and Niger-Congo language families. To facilitate a discussion of linguistically significant E-M96 mutations, presentation of downstream variation within this haplogroup has been divided into four different "clusters" each with a color designation: orange, yellow, blue, red, and green. This step was taken because the internal phylogeny of haplogroup E-M96 is extraordinarily complex compared to the other Y-chromosome haplogroups. The reader is now invited to review **Supplementary Figures 5.1**, **5.2**, **5.3**, and **5.4** which depicts the phylogenetic relationships among the linguistically informative E-M96 clusters.

Section 2. The Evolutionary History of Haplogroup E-M96.

The reader is invited to review **Supplementary Figure 1.1** from the first chapter. Haplogroups E-M96 and D-M174 evolved from the DE-M145 paragroup. According to Poznik et al. (2016), this occurred roughly 67 thousand years ago. Furthermore, as explained previously in Chapter 4, the DE-M145 mutation and its downstream variants, D-M174 and E-M96, have a unique Alu insertion (YAP) polymorphism.

Among the geneticists (e.g., Abu-Amero et al. 2009) some support the position that haplogroups E-M96 and D-M174 evolved outside of Africa in the Middle East. Haplogroup D-M174 then expanded into East Asia about 50 thousand years ago (see Chapter 4 for more details). Haplogroup E-M96, on the other hand, "back-migrated" to Africa by around 56 thousand years ago (Poznik et al. 2016), which correlates well with the beginning of Marine Isotope Stage 3. Interestingly, some argue (e.g., ISOGG 2017) that haplogroup E-M96 evolved in Africa because most of the E-M96 sub-haplogroups evolved on the African continent. However, as suggested by Poznik et al. (2016), a more "parsimonious interpretation" of the data places the origins of E-M96 in the Middle East because otherwise haplogroup D-M174 as well as the C-M130 and FR-M89 paragroups, would have been part of the out of African migration, which is inconsistent with the genetic evidence and archaeological record as detailed previously in Chapter 4.

As noted above, the evolutionary history of the E-M96 haplogroup is incredibly complex. Taking this a step further, deciphering evolutionary relationships within the haplogroup is challenging because over the last 50 thousand years, extensive geneflow has occurred between northern Africa and the Middle East. For example, the E-M96 haplogroup represents *back-to-Africa*; E1b-M123 represents *out-of-Africa-again*; and E1b-M34 *represents back-to-Africa-once-again*.

Geneflow has occurred between northern Africa and the Middle East over the last 50 thousand years is supported by the archaeological record. Bar-Yosef (1987) compared similarities between stone tool industries of the Emiran cultural tradition of the Levant and the Dabban tradition of Cyrenaica in northeastern Libya. Based on these data, he proposed the flow of people and technology between North Africa and the Levant during Marine Isotope Stage 3. Additional support for this conclusion comes from two studies. In 2011, Rebollo et al. reported that the Emiran culture evolved in the Levant roughly 45 thousand years. This estimate was derived from radiocarbon data obtained from organic material at the Kebara cave at Mt. Carmel in Israel. A similar estimate was obtained for the Dabban culture from data obtained at the Haua Fteah cave in Cyrenaica (Douka et al. 2014). Interestingly, the Dabban and Emiran cultures evolved at a transition point in the archeological record, one that marks the end of the Middle Stone Age in Africa and the beginning of the Late Stone Age.

Besides the archeological record, genetic evidence of admixture between Neandertal and modern human can determine the timing of a back-to-Africa migration during Marine Isotope Stage 3 (see Garcea 2016 for more details). Support for this idea is provided by Chen et al. (2020). In this study researchers take the position that back-migration to Africa explains the reduced amount Neandertal genes found among the contemporary populations of the continent. According to Sankararaman et al. (2012), interbreeding between Neandertals and modern humans "most likely" occurred 47 to 65 thousand years ago. As such, admixture data suggest that the reduced presence of Neanderthal DNA among contemporary Africans resulted from the admixture between those who remained in Africa during Marine Isotope Stage 5 and those who back migrated to the continent at the beginning of Marine Isotope Stage 3.

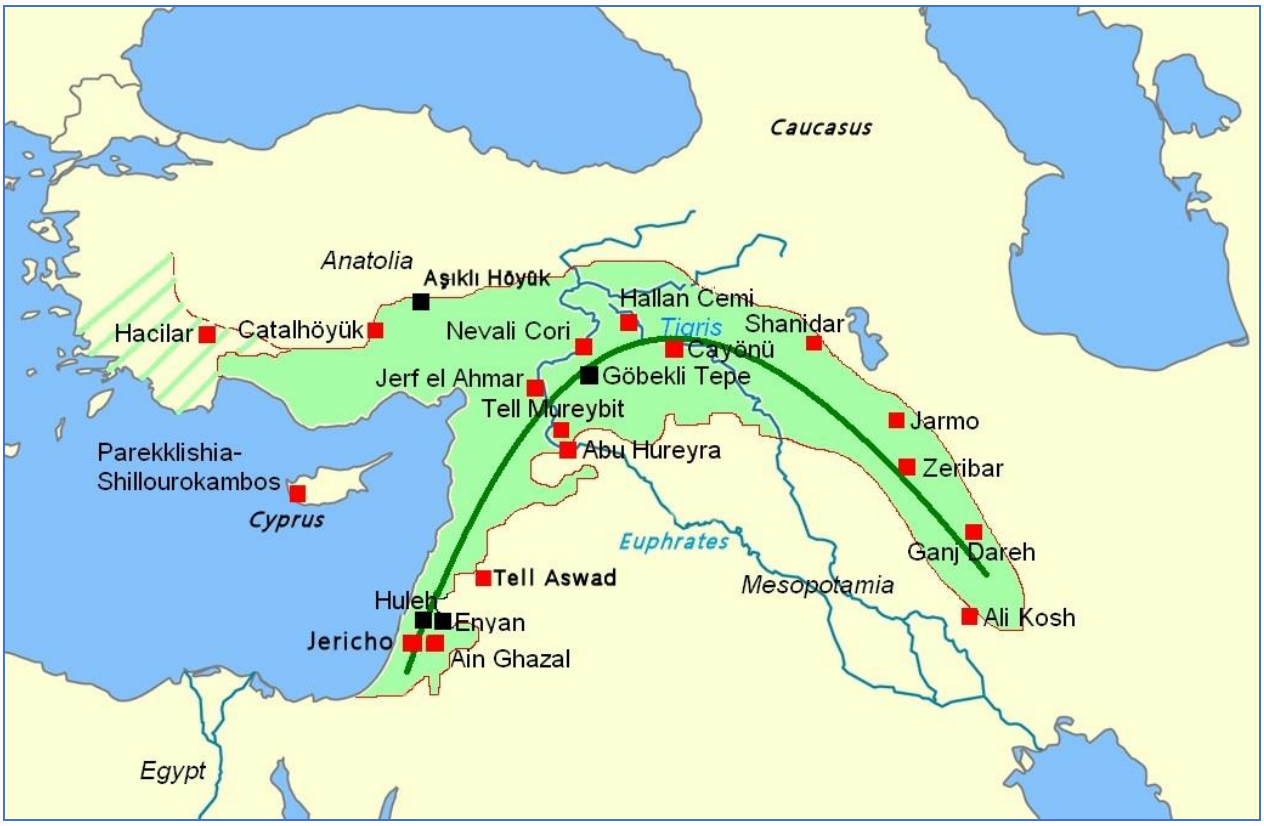
As noted previously in Chapter 4, modern humans inherited autosomal DNA from the Neanderthals but not their Y-chromosome. Nevertheless, from a Y-chromosome perspective, haplogroups A and B-M60 are genetic relics of human that remained in Africa, and haplogroup E-M96 is a relic of the back-migration. Then, in Africa, the initial diversification of haplogroup E-M96 began with the evolution of the E1-P147 and E2-M75 mutations about 50 thousand years (Poznik et al. 2016).

Section 3. The Origins of Agriculture in Southwest Asia.

The 2005 book *First Farmers: The Origins of Agricultural Societies* by Peter Bellwood provides an excellent resource for linguists who wish to explore the worldwide correlation between the origins of agriculture and the expansion of languages. In chapter three of the book (pp. 44-66) he explores the origins of agriculture in Southwest Asia focusing on a region often identified in the literature as the "Fertile Crescent." This region encompasses parts of contemporary Egypt, Israel, Jordan, Lebanon, Turkey, Syria, Iraq, and Iran. The transition to agriculture in the Fertile Crescent was facilitated by the domestication of cereals such as wheat and barley, and legumes such as chickpeas and lentils. Additionally, the agricultural transition in the Fertile Crescent involved the domestication of goats and sheep.

Prior to the adoption of agriculture in Southwest Asia, and elsewhere in the world for that matter, the human tribe employed hunter-gatherer techniques to survive. The evolution of agricultural in Southwest Asia generally follows a series of cultural transitions that began with the Natufians, followed by the Pre-Pottery Neolithic A and Pre-Pottery Neolithic B traditions, and then finally the development of pottery itself. The Natufians stand as an important cultural transition because they were the last hunter-gatherers in the Middle East. According to Bellwood (2005), about 14.5 thousand years ago the Natufians appeared near the Sea of Galilee in what is now present-day Israel. According to Bar-Josef (1998), the Natufians hunted gazelles and other animals. Moreover, and more significantly, he reports that they "practiced intensive and extensive harvesting of wild cereals" that grew abundantly in the region at the time. According to Bellwood (2005), because of this abundant supply of food, he Natufians were able to construct semi-permanent settlements, something unusual for hunter-gatherers. Such cultures are generally nomadic. The Natufians thrived until about 13 thousand years ago when the Younger Dryas cold snap suddenly appeared. For a period of about 700 years global temperatures sank considerably. Weather conditions in Southwest Asia became cold and arid, and with that the abundant supply of wild cereals disappeared. The Natufians became nomads once again and ultimately faded from the archaeological record (see Blockley and Pinhasi 2011).

Figure 5.1. The Fertile Crescent 7500 BC. Source: Wikipedia and Bjoertvedt. Notes: Red squares represent Neolithic sites, and pre-Neolithic sites are represented by black squares.



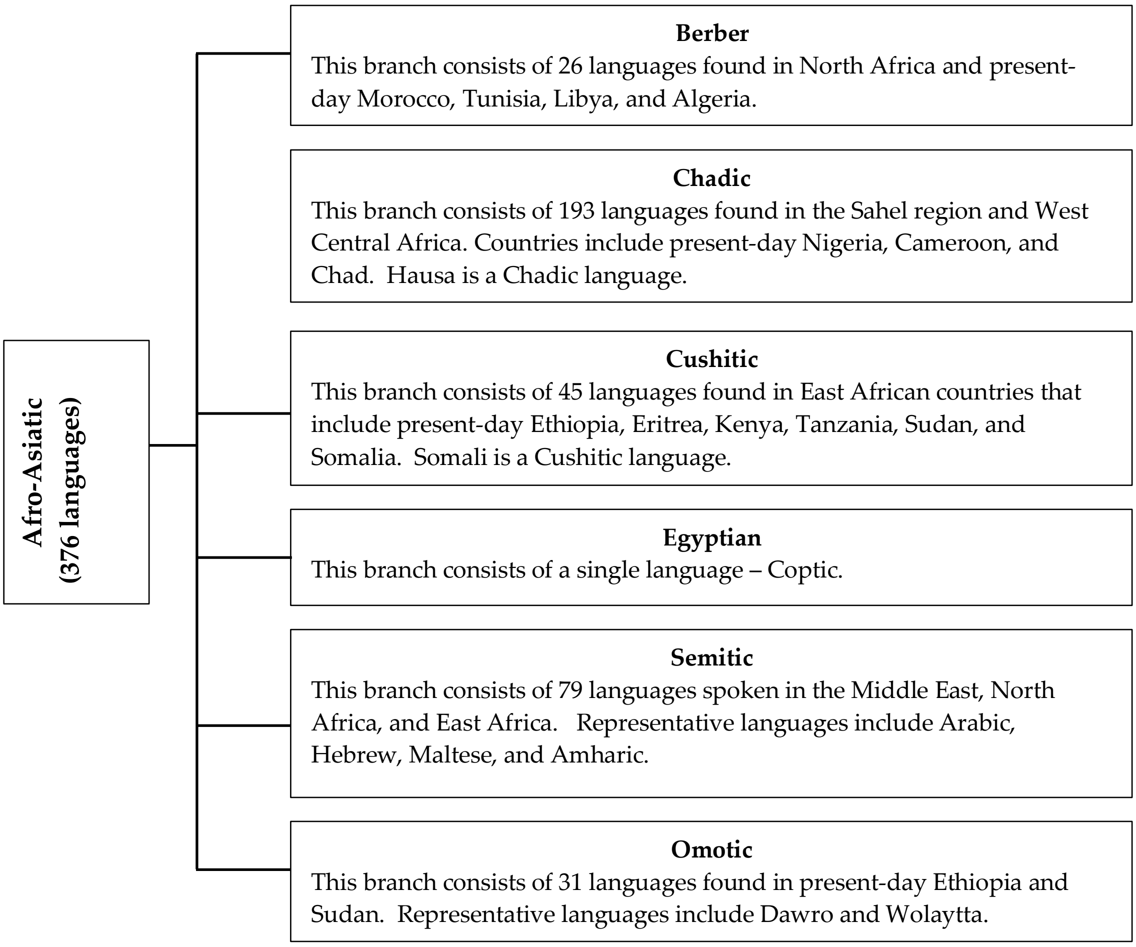
Then almost as suddenly as it began, the Younger Dryas ended, and warmer weather returned. This created ideal climatic conditions that produced, once again, what must have been a seemingly inexhaustible abundance of wild cereals (Bar-Yosef 1998; Bellwood 2005). Amid this abundance, for reasons not entirely clear, a significant human innovation occurred. People began to domesticate the wild cereals and legumes that their Natufian ancestors had previously gathered. The Pre-Pottery Neolithic A culture stands as the initial Southwest Asian culture that embraced this new development. They and their descendants thrived and by around 10.5 thousand years ago large farming settlements appeared such as the one at Abu Hureyra in northern Syria. This development signaled the evolution of another cultural transition in the region, the Pre-Pottery-Neolithic B culture. One of the significant innovations that occurred during this period was the development of pastoralism, the herding of goats and sheep, formerly wild animals that people had managed to domesticate.

About 9,000 years ago the development of pottery ushered in a new cultural transition in Southwest Asia. This development allowed people to cook their food more efficiently and facilitated the storage of grain after harvesting. Around this time the climate in Southwest Asia also became more arid. According to Bellwood (2005) this change in climate was accompanied by deforestation and less productive soil due to over-farming. These conditions caused many people in Southwest Asia to abandon sedentary crop agriculture. Instead of cultivating crops, some turned to sheep and goat herding as a food source. By around 6,400 years ago some of these Southwest Asian pastoralists herded their goats and sheep out of the region into Egypt (Kuper and Kroepelin 2006).

Section 4. The Origins of Afro-Asiatic.

The Afro-Asiatic language family contains 376 languages (Ethnologue 2017). These languages are distributed throughout the Middle East as well as North Africa, East Africa, and the Sahel. At this point the reader is directed to Figure 2.1 from Chapter 2 which displays the distribution of Afro-Asiatic languages within Africa. Additionally, Figure 5.2 (below) lists the language branches of the Afro-Asiatic family and their contemporary geographic distribution. As shown by the figure, Afro-Asiatic is subdivided into six main branches: Egyptian, Semitic, Chadic, Cushitic, Omotic, and Berber. As inferred by the present-day distribution of these six main branches, Semitic evolved in Southwest Asia, whereas Egyptian, Chadic, Cushitic, Omotic, and Berber evolved in Africa.

Figure 5.2. Overview of Afro-Asiatic and it Main Branches. Source: Ethnologue 2017.



Afro-Asiatic (376 languages)

Berber
This branch consists of 26 languages found in North Africa and present-day Morocco, Tunisia, Libya, and Algeria.

Chadic
This branch consists of 193 languages found in the Sahel region and West Central Africa. Countries include present-day Nigeria, Cameroon, and Chad. Hausa is a Chadic language.

Cushitic
This branch consists of 45 languages found in East African countries that include present-day Ethiopia, Eritrea, Kenya, Tanzania, Sudan, and Somalia. Somali is a Cushitic language.

Egyptian
This branch consists of a single language – Coptic.

Semitic
This branch consists of 79 languages spoken in the Middle East, North Africa, and East Africa. Representative languages include Arabic, Hebrew, Maltese, and Amharic.

Omotic
This branch consists of 31 languages found in present-day Ethiopia and Sudan. Representative languages include Dawro and Wolaytta.

Long-standing opinion among linguists (e.g., Ehret 2004) places the prehistoric origins of Afro-Asiatic languages somewhere in East Africa. This opinion follows the idea that the putative homeland of a language lies where its greatest diversification is found (e.g., Hetzron 2009). However, Bellwood (2005: 207-210), based on his interpretation of the archaeological data, suggests that Afro-Asiatic languages initially evolved in Southwest Asia and co-expanded out of this region with the spread of agriculture. Interestingly, linguistic data may also support this model of Afro-Asiatic origins. Using linguistic reconstructions, Militarev (2002) presents a proto-Afro-Asiatic lexicon of farming terminology. Based on the reconstructions, he suggests that the Natufians, agriculture, and Afro-Asiatic co-evolved in Southwest Asia.

Section 5. The Green Cluster Mutations.

5.1. Overview.

Again, the reader is reminded that the evolutionary history of E-M96 becomes complicated because of prehistoric pendulum-like geneflow between North Africa and the Middle East. Additionally, it should be noted that *E1b-M35* is upstream, and *E1b-M34* is downstream. At this point the reader is directed to **Supplementary Figure 5.1**. The E1b-M35 mutation evolved roughly 44 thousand years ago (Poznik et al. 2016). This probably occurred in Africa. Later, the green and red clusters mutations evolved from the E1b-M35 mutation. The green cluster mutation is defined by the E1b-Z827 mutation. Linguistically informative E1b-Z827 mutations are E1b-M34, E1b-M81, and E1b-M293.

5.2. E1b-M34 and Afro-Asiatic.

The E1b-M34 mutation has a wide distribution, currently found in populations of Mediterranean Europe, southeastern Europe, the Middle East, North Africa, and East Africa (see **Supplementary Table 5.1**). For linguists, the E1b-M34 mutation is significant marker. Ancient and contemporary Y-chromosome data suggest that this marker was carried from the Middle East to North Africa by goat and sheep herders about 6,000 years ago. This Neolithic migration, as the reader may recall from Section 4 (above), appears to have carried Afro-Asiatic languages into Africa.

Analysis of contemporary genetic data suggest that E1b-M34 evolved in the Middle East (Cruciani et al. 2004; Cadenas et al. 2008). Support for this conclusion is also provided by ancient Y-chromosome data (see **Supplementary Table 5.2**). Turning now to the ancient data, upstream from the E1b-M34 marker is E1b-PF1961 mutation. A study from 2016 (Lazaridis et al.) was able to extract three ancient DNA samples from a Natufian archaeological site in Israel. As the reader may recall from Section 3 (above), the Natufians were the last hunter-gathers of Southwest Asia. Two of the samples belong to E1b-PF1961. These ancient Y-chromosome data suggest that E1b-PF1961 migrated out of Africa and underwent diversification in the Middle East before back-migrating to Africa. It is difficult to determine when the E1b-PF1961 mutation arrived in the Middle East. Trombetta et al. (2015) suggest sometime within the last 25 thousand years ago. Several studies suggest this "out-of-Africa" migration probably followed the Nile River as it would have been an ideal corridor for human expansions (see Cruciani et al 2004; Luis et al. 2004; Cruciani et al. 2007; Cadenas et al. 2008).

5.3. E1b-M81, Afro-Asiatic, and Berbers.

Another linguistically significant green cluster mutation E1b-M81. Its position within the haplogroup E-M96 phylogeny (see **Supplementary Figure 5.3**) and its contemporary distribution (see **Supplementary Table 5.3**) suggest that E1b-M81 arose somewhere in northwestern Africa. In their 2004 analysis of the contemporary E1b-M81 data, Arredi et al. identify the E1b-M81 mutation as a Neolithic marker. They suggest that goat and sheep pastoralism from Southwest Asia produced a "demic diffusion" of this mutation across North Africa. The term "demic diffusion" describes a scenario where a group adopts agriculture. The group then expands into an uninhabited region. Rapid population growth follows because agriculture supports far more people per square kilometer than hunter-gatherer food economies. Y-chromosome mutations frequently ride the coattails of these expansions and produce a clinal frequency pattern that increases over distance. This explains why the E-M81 mutation has a low frequency in Egypt and a high frequency in Morocco. It should be noted that an ancient DNA study provides additional support that links E1b-M81 with the North African Neolithic. Fregel et al. (2018) reports the discovery of the E1b-M81 mutation among Neolithic remains from Morocco (see, also, **Supplementary Table 5.1**).

The E1b-M81 mutation consistently attains a high frequency among Berber populations in North Africa (e.g., Bosch et al. 2001; Ennafaa et al. 2011; Fadhlaoui-Zid, et al. 2011; Trombetta et al. 2015; and **Supplementary Table 5.3**). As such, this marker becomes not only the genetic signature of the North African Neolithic but also the genetic signature of Berber languages, a branch of the Afro-Asiatic language family. In addition to the Berber people, the Tuareg people of the Sahara Desert also speak Berber languages. The linguistic evidence along with the elevated frequency of E-M81 among these people suggest that they descended from the same ancestral as the Berbers (for more details, see Pereira et al. 2010; Ottoni et al. 2011).

The researcher Roger Blench (2014) posted an interesting paper on his website that presents an anthropological and linguistic perspective of the Berber people and language. According to the paper, a comparison of grammar suggests that Semitic is the closest Afro-Asiatic branch to Berber. Blench also suggests that the Berber branch split from Afro-Asiatic language family around 6,500 years ago. However, he acknowledges that time depth of this magnitude seems inconsistent with close linguistic similarities as found among the 26 contemporary Berber languages. Bench argues that a leveling of linguistic differences among the Berber languages occurred about 2,000 ago. According to the researcher, support for this conclusion comes from Neo-Punic and Latin lexical borrowings found in contemporary Berber languages. Blench further suggests that the expansion of the Roman Empire into North Africa created a need for a *lingua franca* among the Berbers. By this time the Berbers rode camels, and this brought an opportunity to trade with the Romans, especially along their southern frontier in North Africa, the so-called "limes." Thus, a *lingua franca* among the Berbers facilitated trade with the Romans. According to Bench, the adoption of a common trade language among the Berbers ultimately leveled linguistic diversity among this people. Blench further writes that the influence of the Berbers in North Africa later diminished after the spread of Islam throughout the region.

A discussion of expansion of Afro-Asiatic languages across North Africa continues in Chapter 11 and the discussion of haplogroup J-M304. A variant of this haplogroup, the J1-M267 mutation, which has origins in Southwest Asia, co-expanded across North Africa with E-M81.

5.4. E1b-M293, Nilo-Saharan, Khoisan, and Niger-Congo.

A final linguistically significant green cluster mutation is E1b-M293. Trombetta et al. (2015) estimate that this mutation evolved about 3.5 thousand years ago. In their 2008 study, Henn et al. suggest that the mutation evolved in Tanzania among the Datooga people. This population speaks a Nilo-Saharan language. However, as demonstrated by the data in **Supplementary Table 5.4**, the E1b-M293 mutation not only attains a significant frequency among Nilo-Saharan populations such as the Datooga, Maasai, and Turkana, but also among populations that speak Afro-Asiatic, Niger-Congo, and Khoisan languages. Additionally, E1b-M293 is found among the Sandawe and Hadza, two populations that speak a language classified as an isolate. The Henn et al. (2008) study suggests that this distribution pattern follows a southward expansion of East African cattle pastoralism that began about 2,000 years ago.

Surprisingly, a 2018 study (Bajić et al.) reports no correlation between pastoralism and the E1b-M293 mutation among contemporary populations. However, this conclusion seems problematic for two reasons. First, the dataset used by the study is too small in that the researchers only considered E1b-M293 samples from Namibia, Botswana, and Zambia. The E1b-M293 mutation is, in fact, observed in populations across a geographical expanse that stretches from Ethiopia to South Africa (see **Supplementary Table 5.4**). Second, the study focuses on statistical modeling and ignores the archaeological and climatological data. Archaeological and climatological perspectives, in fact, link the evolutionary history of the E1b-M293 mutation to an expansion of East African cattle herders and their interactions with Bantu farmers and Khoisan hunter-gatherers. This observation underscores two salient points for linguists: (1) language variation in Africa has good correlation with subsistence strategy; (2) and language variation in Africa was shaped, in part, by language shift and language maintenance.

Archaeological support for the link between the East African cattle herding tradition and the contemporary distribution of E1b-M293 mutations begins was a discussion of the Nabta Playa and Bir Kiseiba archeological sites in southern Egypt. Here, cattle herding became part of the subsistence strategy in Africa around 8,000 years ago (e.g., Brass 2018). Interestingly, anthropologists are divided on the origins of domesticated cattle in Africa. Some researchers favor the domestication of wild African bovines. Others suggest that they were initially domesticated in the Middle East and brought to Africa as part of the Southwest Asian agricultural package (see Brass 2018 for more details). On the other hand, a study from 2014 (Decker et al.) suggests, based on genetic data, that domesticated African cattle are a hybrid of domesticated Middle Eastern cattle and wild African aurochs. For linguists, this controversy raises an interesting question: Were domesticated cattle brought to Africa by Afro-Asiatic speakers?

The expansion of cattle herding from northern Africa to southern Africa is linked to climate change and prevalence of the tsetse fly. This blood sucking insect transmits the so-called "sleeping sickness" while feeding. Domesticated cattle are especially vulnerable. A 1992 paper (Smith) explains that tsetse flies only thrive in regions that receive at least 500 millimeters of rainfall per year. Otherwise, they cannot breed. The same paper times the southern advance of cattle herding from North to East Africa at around four thousand years ago. According to the researcher, this correlates well with the end of the last Saharan humid phase and the southward retreat of the intertropical transition zone. He explains that cattle pastoralism in Africa represents, in effect, a human adaptation to grassland environments that appeared at the result of this change in climate. With increased aridity, grassland expanded and the tsetse because far less prevalent. The absence of tsetse and the expansion of grassland, in turn, ultimately made cattle pastoralism more productive for people in East Africa.

Around 2,000 years ago the Elmentaitan culture was well-established on the western plains of Kenya. This signaled the success of cattle pastoralism in East Africa. By this time climate change in Africa created a tsetse-free corridor that facilitated a secondary expansion of cattle pastoralism into southern Africa (Chritz et al. 2015). This corridor ran parallel to the western coast of Africa (see Smith 1992: Figure 1). As cattle pastoralism expanded southwards from East Africa, another independent agricultural expansion occurred further west. Bantu farmers and cereal cultivation pushed southwards through the central African rainforest. According to Lander and Russell (2018), cattle pastoralism arrived in southern Africa slightly ahead of the Bantus. The Khoisan people, of course, were already in the region. Eventually a collision of all three cultural traditions occurred. According to the archaeological record compiled by Lander and Russell (2018), interaction between the three traditions was complex. Cattle herders traded meat, Bantus traded grain, and the Khoisan provided labor. Their data also suggest that language variation in southern Africa was shaped by assimilation, language shift, and language maintenance. If the cattle pastoralists spoke a Nilo-Saharan language, which is consistent with the genetic and linguistic evidence, the absence of this language family in southern African can be explained by the assimilation of cattle herders into Bantu and Khoisan populations.

Section 6. Red Cluster Mutations.

6.1. Overview.

As noted previously in the discussion of green cluster mutations, haplogroup E-M35 evolved about 44 thousand years ago in Africa. Green cluster mutations are variants of the E1b-Z827 marker (**Supplementary Figure 5.2**) and red cluster mutation are variants of the E1b-M78 marker (see **Supplementary Figure 5.3**). Within the red cluster, three mutations are linguistically informative: E1b-V32, E1b-V13, and E1b-V22. The contemporary distribution of these mutations and ancient Y-chromosome data suggest extensive prehistoric bi-directional geneflow between North Africa and the

Middle East (see **Supplementary Table 5.2**, **5.5**, **5.6**, and **5.7**). The ancient Y-chromosome data was reported by Loosdrecht et al. (2018). The study identified the E1b-M78 mutation among several remains from an archaeological site in Morocco. These remains were dated to about 14.5 thousand years ago, and as such, place the E1b-M78 in North Africa during the Late Paleolithic.

6.2. E1b-V13, E1b-V22, and Afro-Asiatic.

Focusing now on E1b-V13, a red cluster mutation, this is the only haplogroup E-M96 variant that attains a significant frequency in Europe. As shown by **Supplementary Table 5.5**, E-V13 attains a significant frequency among the populations of the Balkans and in Greece. More moderate frequencies are observed elsewhere in Europe, such as among the Italians and the Hungarians. Several studies suggest that E1b-V13 entered Europe during the Mesolithic (Battaglia et al. 2009; Regueiro et al. 2012; Karachanak et al. 2013). This would suggest that the expansion of E1b-V13 into Europe followed the disintegration of the Natufian culture during the Younger Dryas (see discussion in Section 3 above). However, another study (Trombetta et al. 2015) suggests that E-V13 evolved around 8,000 years ago. Here, researchers favor a Neolithic or latter arrival of the mutation in Europe. Moreover, ancient DNA from archeological sites place E1b-V13 in East and West Europe during the Neolithic (see **Supplementary Table 5.2**). These data from Europe raise an interesting question. Were proto-Afro-Asiatic languages part of the linguistic inventory of prehistoric Europe? This possibility was proposed by Vennemann (2000) and Mailhammer (2007).

Focusing now on E1b-V22 mutation, it is difficult to pinpoint where this marker evolved. According to Cruciani et al. (2007), the E-V22 mutation evolved about 10 thousand years ago. This dating result would support the position taken by Hassan et al. (2008), that E1b-V22 is a potential genetic relic of the desertification of the Sahara. However, this position seems inconsistent with the linguistic data in that the mutation fails to exhibit a strong frequency among Nilo-Saharan populations (see **Supplementary Table 5.6**). Rather, the mutation exhibits a strong frequency among Semitic and Cushitic population in North and East Africa. Thus, based on the frequency and linguistic data, E1b-V22 may well represent a Neolithic back-to-Africa migration of farmers or pastoralists that spoke a proto-Afro-Asiatic language. Support for this position stems from its close phylogenic relationship with the E1b-V13 mutation (see **Supplementary Figure 5.3**). If E1b-V13 evolved in the Middle Eastern, E1b-V22 can also trace its origins to the same region. Taking this a step further, E1b-V22 may well have co-migrated into North Africa with the "green cluster" E1b-M34 mutation that was described previously in Section 5.2.

6.3. E1b-V32, Nilo-Saharan, and Afro-Asiatic.

As the reader may recall from Chapter 2 and the discussion of the A1b-M13 mutation, about ten thousand years ago Holocene climate change transformed the Sahara Desert into a savannah type ecosystem complete with rivers and lakes. Then about 7,000 years ago the rain ceased, and suddenly the Sahara became, once again, a desert. As result of the so-called "desertification" of the Sahara, people either congregated along the Nile River in Egypt, or alternatively, moved with their herds of cattle, goats and sheep into the Sudan and East Africa (e.g., Kuper and Kröpelin 2006). Those that settled along the Nile eventually adopted sedentary agriculture and cultivated crops that came from the Levant. The pastoralists, on the other hand, herded sheep, goats, and cattle.

The E1b-V32 marker stands as a strong genetic relic of "desertification" of the Sahara. Cruciani et al. (2007) suggest that the mutation evolved roughly 8,000 ago in northeastern Africa. According to the same study, E1b-V32 currently represents 82 percent of E-M78 (or red cluster) variation in East Africa. As shown by **Supplementary Table 5.7**, the E1b-V32 mutation attains a significant frequency among Nilo-Saharan and Afro-Asiatic populations in East Africa. Thus, the marker potentially represents a population expansion among Nilo-Saharan cattle herders in East Africa at the end of the last humid phase. Assimilation and language shift then explain the presence of the mutation among the Afro-Asiatic populations of the region.

Section 7. Nilo-Saharan and the E2a-M41 Yellow Cluster Mutation.

The reader is directed to **Supplementary Figure 5.1** and the E2a-M41 "yellow cluster" mutation. Little is known about this mutation, including when it diverged from E2-M75. Most of the frequency data comes from a 2010 study (Gomes et al.) and 118 samples taken from three different populations in Uganda: the Dodoth, Jie, and Karimojong. These groups speak Ng'arkarimojong, a Nilo-Saharan language. Overall, the E2a-M41 mutation attains a modest frequency of 11 percent. In their 2005 study, Wood et al. report that this mutation attains a frequency of 67 percent among the Alur people of the Democratic Republic of the Congo, a population that also speaks a Nilo-Saharan language. However, the sample size was small (nine men) and ascertainment bias may well have skewed the actual frequency. The only other African population in which E-M41 attains a significant frequency is the Hema ethnic group in the Democratic Republic of the Congo where the mutation is reported in 39 percent of the men (Wood et al. 2005). This population speaks a Bantoid language.

Gomes et al. (2010) suggest that the E2a-M41 mutation represents a potential marker for understanding the genetic history of Nilo-Saharan speaking populations in East Africa. Indeed, the data suggest that the potential genetic relics of pre-agricultural Nilo-Saharan speaking populations in Africa include not only the E2a-M41 "yellow cluster" mutation, but also the E1a-M33 "orange cluster" mutation (see Section 8 below).

Section 8. Nilo-Saharan, Afro-Asiatic, Niger-Congo and Orange E1a-M33.

According to data from Poznik et al. (2016), E1-P147 diverged from the E-M96 main haplogroup about 50 thousand years ago. Shortly thereafter, about 48 thousand years ago, E1a-M33 diverged from E1-P147. These dating estimates, along with its position within the E-M96 main haplogroup phylogeny (see **Supplementary Figure 5.1**), reflect that the E1a-M33 orange cluster mutation evolved shortly after the initial back-to-Africa migration, which occurred by around 50 thousand years ago. As such, E1a-M33 represents a comparatively ancient mutation that traces its origins close to the initial diversification of E-M96 variation in Africa. As shown by **Supplementary Table 5.8**, the geographic distribution of E1a-M33 populations is rather interesting as these populations are found in the Sahel region of Africa, a transition region between the southern border of the Sahara Desert and Central African rainforest. Moreover, this region represents the putative homeland of Nilo-Saharan languages (see discussion in Chapter 2).

Within the Sahel region, **Supplementary Table 5.8** reflects that E1a-M33 attains a moderate frequency among Nilo-Saharan, Niger-Congo and Afro-Asiatic populations. Interestingly, the presence of E1a-M33 among Mande speakers and non-Bantoid Atlantic-Congo speakers confirms what the linguistic evidence suggests: Mande and non-Bantoid Atlantic-Congo populations are geographically closer to the putative Niger-Congo homeland in West-Central Africa. In their survey of populations in West-Central Africa, Filippo et al. (2011) and Barbieri et al. (2012) analyzed genetic diversity among Mande and non-Bantoid Atlantic-Congo speakers and reached a similar conclusion. This explains why Mande and non-Bantoid Atlantic-Congo populations tend to have the orange cluster E1a-M33 mutation as well as undefined older blue cluster mutations that are downstream from E1b-M2. The Bantus, on the other hand, tend to have, almost exclusively, blue cluster E1b-U174 and E1b-U175 variants. Taking this a step further, the Bantu expansion conforms to a demic diffusion model, one that also explains the expansion of Berber languages across North Africa (see Section 5.3 above).

Turning now to Afro-Asiatic, the presence of the E1a-M33 mutation among Chadic speaking populations, such as the Kotoko and Masa of Cameroon (see Bučková et al 2013), and the Hausa of

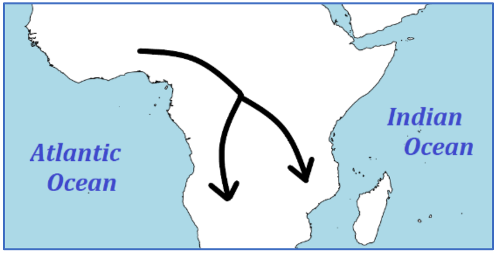
Sudan (see Hassan et al. 2008), also appears to be significant. Chadic populations also have a significant frequency of the R1b-V88 mutation (see Chapter 18). These data suggest that language shift may well explain the contemporary distribution of Afro-Asiatic language in Africa.

Section 9. Bantu Farmers, Pygmies, Khoisan, and Blue Cluster Mutations.

As explained previously in Chapter 3, the linguistic prehistory of Bantoid languages is linked to a southward expansion of farmers from West-Central Africa about four thousand years ago. Additionally, as noted in Chapter 3, the B2a-M150 mutation represents a genetic relic of this expansion. Like the B2a-M150 mutation, the E1b-U174 and E1b-U175 mutations have been identified as especially strong genetic relics of the Bantu expansion (e.g., Filippo et al. 2011; Montano et al. 2011; Barbieri et al. 2012; Rowold J. et al. 2016). Frequency data for both mutations support this position (see **Supplementary Table 5.9** and **Supplementary Table 5.10**). Additional support comes from dating estimates. According to Filippo et al. (2011), the E1b-U175 mutation evolved in West-Central Africa about 5,000 years ago, and E1b-U174 evolved in the same region about 4,000 years ago. These dating estimates agree with the timing of the Bantu expansion as provided by climatological and anthropological perspectives (see discussion in Chapter 3).

The position of the E1b-U174 and E1b-U175 mutations within the haplogroup E-M96 phylogeny is diagramed in **Supplementary Figures 5.1.** and **5.4**. As shown by **Supplementary Figure 5.1**., the E1b-P2 mutation unites the blue, red, and green clusters. The blue cluster is defined by the E1b-V38 mutation. As shown by **Supplementary Figure 5.4**, linguistically informative variants of the E1b-V38 "blue cluster" mutation are E1b-U174 and E1b-U175.

Figure 5.3. The Bantu Expansion.



The discussion of B2a-M150 and B2b-M112 mutations in Chapter 3 suggests that these mutations can potentially measure male geneflow between Bantu farmers and African foragers. These data, in turn, point to factors that produced language shift among the Pygmies, and language maintenance among the Khoisan. The E1b-U174 and E1b-U175 mutations are also a potential source of data for exploring this question. As shown by **Supplementary Tables 5.11** and **5.12**, the available Y-chromosome data suggest substantial unidirectional geneflow from Bantus into Khoisan and Pygmy groups. However, caution is urged against making big conclusions with little data. The amount of Y-chromosome data for Sub-Saharan Africa is still limited.

Section 10. Conclusions for Haplogroup E-M96.

Several downstream variants of the E-M96 haplogroup help to decipher the origins and expansion of languages on the African continent. The red cluster E1b-V22 and green cluster E1b-M34 mutations represent Afro-Asiatic agriculturalist that entered North and East Africa during the Neolithic. Proto-Berber and the green cluster E-M81 mutation co-expanded across North-Africa. The orange cluster E1a-M33 and yellow cluster E2a-M41 mutations are genetic relics of pre-agricultural Africa populations that predate the evolution of the Niger-Congo and Nilo-Saharan language families. E1b-V32 from the red cluster represents a genetic relic of cattle pastoralism among the Nilo-Saharan populations of East Africa. The green cluster E1b-M293 mutation represents a later southward expansion of pastoralism from East Africa. The E1b-U174 and E1b-U175 mutations carry the Bantu expansion southwards from West-Central Africa. From a Y-chromosome, these linguistically informative mutations define factors that have shaped contemporary language variations in Africa, which include climate change, agricultural expansions, language shift, and language maintenance.

Chapter 6: Haplogroup C1-F3393.

Section 1. Contemporary Distribution of C1-F3393.

The designation of C1-F3393 as a haplogroup is nonstandard (see the Y-chromosome tree that is posted by the International Society for Genetic Genealogy). An explanation follows in Section 2.1 (below).

The reader is now invited to examine **Supplementary Table 6.1** which provides frequency data for the C1-F3393 haplogroup from a regional perspective. As shown by the table, haplogroup C1-F3393 attains a significant frequency among the contemporary populations of Island Southeast Asia and Oceania. Additionally, the haplogroup is a significant mutation among Aboriginal Australians. Finally, the haplogroup is observed among the populations of South and East Asia, where it generally attains a frequency of less than 10 percent.

Section 2. Evolutionary History of C1-F3393.

The nomenclature for Y-chromosome mutations was initially standardized in 2002 by the Y-chromosome Commission (see Chapter 1 for more details). Among the measures taken by the commission is the utilization of "haplogroup" to define a unique segment of human Y-chromosome variation. Among the haplogroups that emerged from this standardization was C-M130. However, based on the data that have accumulated over the last twenty years, researcher should relabel the M130 mutation as a higher paragroup, and identify the C1-F3393 and C2-M217 mutations as Y-chromosome haplogroups. As detailed in this present chapter and in Chapter 7, both mutations have evolutionary histories that are quite unique. The contemporary pattern of C1-F3393 mutation resulted from human expansions during Marine Isotope Stage 3, beginning roughly 50 thousand years ago. The contemporary distribution of C2-M217, on the other hand, is a significant haplogroup for defining human expansion after the Last Glacial Maximum, roughly 20 thousand years ago.

At this point the reader may want to identify the position of the DR-M168 mutation within the Y-chromosome tree (see **Supplementary Figure 1.1** from the first chapter). As discussed in Chapter 4, the DR-M168 mutation represents the ancestral mutation for all non-African haplogroups. Dating estimates from Poznik et al. (2016) suggest that this paragroup evolved in the Levant roughly 100 thousand years ago. In this region, roughly 70 thousand years ago, the C-M130 paragroup evolved from the DR-M168 mutation. Diverging from C-M130 are the C1-F3393 and C2-M217 haplogroups. This split occurred about 49 thousand years ago. As shown by **Supplementary Figure 6.1**, C1a-CTS11043 and C1b-F1370 form the two main phylogenetic division with the C1-F3393 haplogroup. Mutations downstream from C1b-F1370 represent human expansion via the "southern route" during Marine Isotope Stage 3. Mutations downstream from C1a-CTS11043 represent expansions via the "northern route" during the same period.

Section 3. Dispersal of C1b-F1370 Mutations via the Southern Route.

3.1. Overview of the Genetic Data.

As previously detailed in Chapter 4, human expansions via the "southern route" explain the successful human colonization of South Asia, East Asia, Island Southeast Asia, and Australia during

Marine Isotope Stage 3, roughly 50 thousand years ago. The available data also suggest that the southern migration began somewhere in the general vicinity the Black Sea. The human tribe then migrated southwards along the Tigris and Euphrates Rivers. From the Persian Gulf, the migration followed the western coastline of India to Sri Lanka and then turned northwards along the Bay of Bengal to Southeast Asia. At this location, C1b-F1370 mutations then expanded southwards through the Malay peninsula into Sunda and Sahul.

As shown by **Supplementary Figure 6.1**, C1b-F1370 splits into C1b-K281 and C1b-B477. Downstream from C1b-K281 are the C1b-M356 (yellow border) and C1b-AM00847 (orange border) mutations. Poznik et al. (2016) suggest that this split occurred about 44 thousand years ago. The C1b-M356 mutation probably evolved in South Asia, the region where the mutation attains its highest frequency among contemporary populations. C1b-AM00847, the phylogenetic "sister" mutation of C1b-M356, probably evolved in East Asia where it is occasionally surfaces among the contemporary populations of the region (see Macholdt et al. 2020; Kutanan et al. 2020).

Downstream from C1b-B477 (see **Supplementary Figure 6.1**) are the C1b-M38 (green border) and C1b-M347 (blue border) mutations. Karmin et al. (2015) suggest that C1b-M38 evolved from C1b-B477 about 24 thousand years ago. According to Mona et al. (2007), this occurred in the northwestern part of New Guinea. C1b-M347, on the other hand, probably evolved in Australia about 44 thousand years ago (see Nagle et al. 2016a).

In published studies only a single informative downstream variant of the C1b-P38 marker has emerged, the C1b-M208 mutation. This marker evolved about 12 thousand years ago in the highlands of western New Guinea (see Delfin et al. 2012; Karmin et al. 2015).

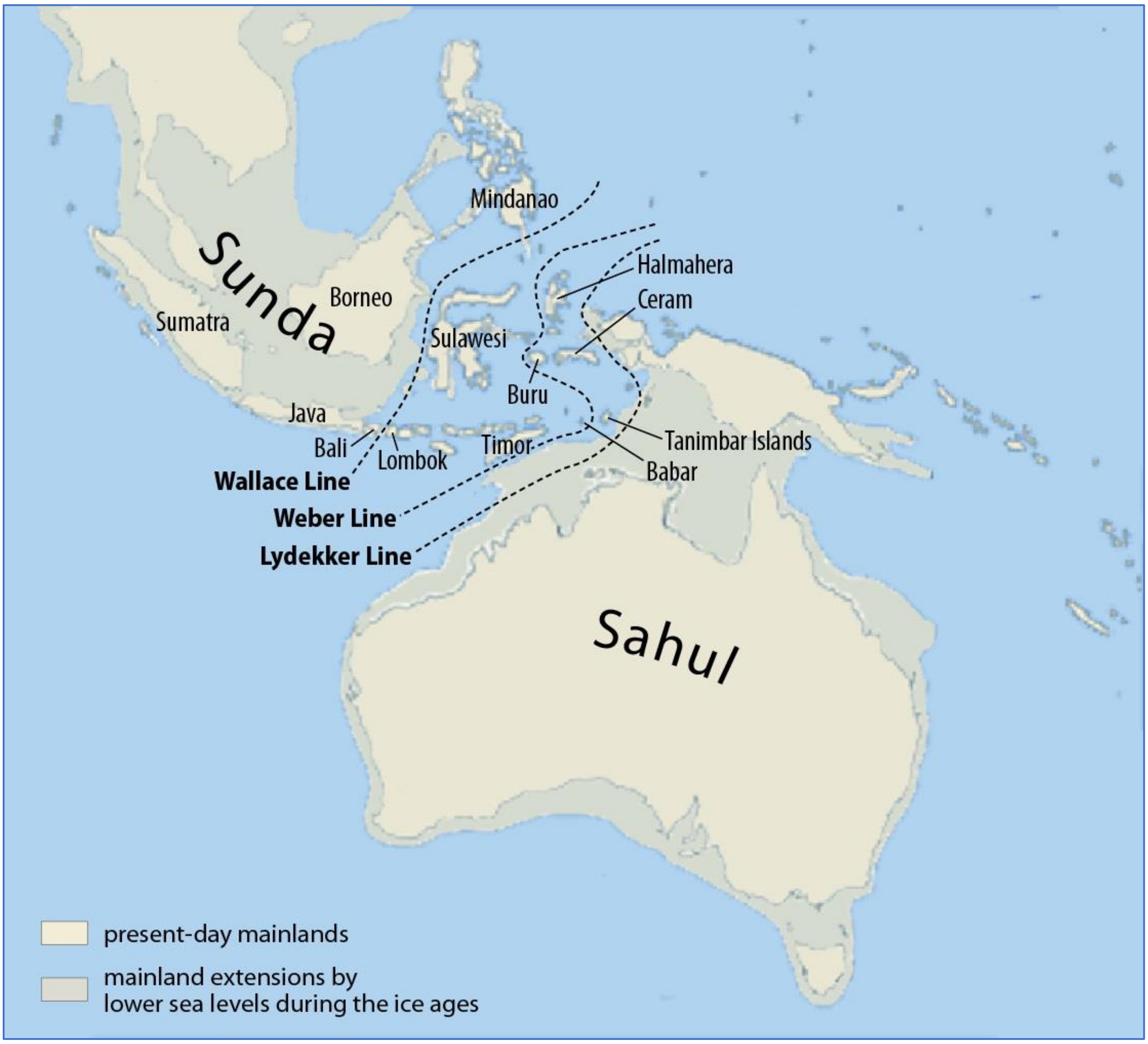
3.2. Archaeological and Climatological Perspectives.

As suggested previously in Chapter 4, more stable weather at the onset of Marine Isotope Stage 3 produced more plentiful marine resources that fueled human expansion along the South Asian coastline. This migration, in turn, facilitated the human colonization of South Asia. The contemporary distribution of C1b-AM00847 and C1b-M356 mutations support this expansion model.

Human migration into Sunda and Sahul from southeastern Asia was facilitated by lower sea levels during the Marine Isotope Stage 3 that emerged as the result of glaciation (see Clark et al. 2009 for more details). Consequently, a large landmass called "Sunda" connected the present-day Malaysian Peninsula with many of the contemporary Indonesian Islands, including Sumatra, Java, Borneo, and Bali. At the same time the Sahul landmass connected Papua New Guinea and Australia (see Figure 6.1 below for additional details). Since the distance between Sunda and Sahul may have been as short as ninety kilometers, the human settlement of Australia may have been accomplished with the use of primitive watercraft (see Allen and O'Connell 2008). This explains, in turn, the contemporary distribution of C1b-M38 mutations in Island Southeast Asia, and C1b-M347 mutations in Australia, with the idea that both mutations represent diversification of the C1b-B477 mutation among geographically isolated populations.

Important fossil remains from human colonization Sunda and Sahul during Marine Isotope Stage 3 include Lake Mungo man from Australia. These remains date to at least 46 thousand years ago (see Bowler et al. 2003). Another important find is the so-called "Deep Skull" at the Niah Cave on the Indonesian Island of Borneo, which date to at least 35 thousand years old (Barker et al. 2007). Additional archeological support comes from a recent paper (Florin et al. 2020). The data consist of charred food remains from Madjedbebe, an archaeological site in Australia. These remains show that modern humans processed food at this location by around 53 thousand years ago.

Figure 6.1. Sunda and Sahul. Source: Wikipedia and Maximilian Dörrbecker (Chumwa).



Section 4. C1b-M356 and Linguistic Diversity in South Asia.

According to Poznik et al. (2016), the C1b-M356 mutation evolved about 48 thousand years ago. Despite the low frequency numbers among contemporary populations (see **Supplementary Table 6.2**), the mutation is still a significant marker for investigating the population history of South Asians. In this region it represents the genetic relic of the southern migration route and the founding population of the region (e.g., Sengupta et al. 2006; Arunkumar et al. 2012; Khurana et al. 2014).

Contemporary South Asia consists of India and Pakistan and is inhabited by over 1.5 billion people (*CIA World Factbook* 2017). Linguistic diversity within the region is incredibly complex: Dravidian; Indo-Aryan, Munda, and Tibeto-Burman. The C1b-M356 mutation represents a starting point for gaining an understanding of this diversity. Within the genome of this region, some mutations, like the C1-M356 marker, represent a Paleolithic component. Other mutations represent more recent migrations during the Mesolithic or Neolithic. Accordingly, the discussion of linguistic diversity in South Asia continues in Chapter 8 with the presentation of haplogroup H-M2713, another Paleolithic component of South Asian populations.

Section 5. C1b-M38, C1b-M208, Papuan, and Austronesian.

5.1. Overview.

At this point it is necessary to explain why **Supplementary Table 6.3** reports data for C1b-M38 without the C1b-M208 mutation (C1b-M38 xM208) and why **Supplementary Table 6.4** provides data for the C1b-M208 mutation. The C1b-M208 mutation is the only informative C1b-M38 downstream variant that has been identified in published sources. C1b-M38 xM208, on the other hand, means that C1b-M38 has additional informative downstream mutations that await discovery. As shown by **Supplementary Table 6.3**, C1b-M38 without C1b-M208 attains a significant frequency in Island Southeast Asia and is essentially absent in Oceania. C1b-M208, on the other hand, attains a significant frequency in Oceania and is essentially absent in Island Southeast Asia.

5.2. Geography.

In order to discuss the prehistory of Papuan and Austronesian languages, it is necessary, at this point, to define geographical terminology. I define Island Southeast Asia as the Philippines, Indonesia east of the Wallace Line, East Timor, and Papua New Guinea. Oceania, on the other hand, consists of a broad expanse of islands in the Pacific Ocean that runs eastwards from the Solomon Islands to Rapa Nui, and southwards from the Hawaiian Islands to New Zealand. It should be noted that this description of the geography is somewhat non-standard and that terms such as Micronesia, Melanesia, and Polynesia are more common. Additionally, the literature sometimes describes the geography as Near Oceania, Remote Oceania, and Australasia. Finally, some regard the Philippines as part of East Asia. While the geographical descriptors used in this chapter might be non-standard, the regional descriptions are necessary in order to facilitate an efficient delivery of the linguistic and genetic data. For additional information, see Figure 6.2 (below).

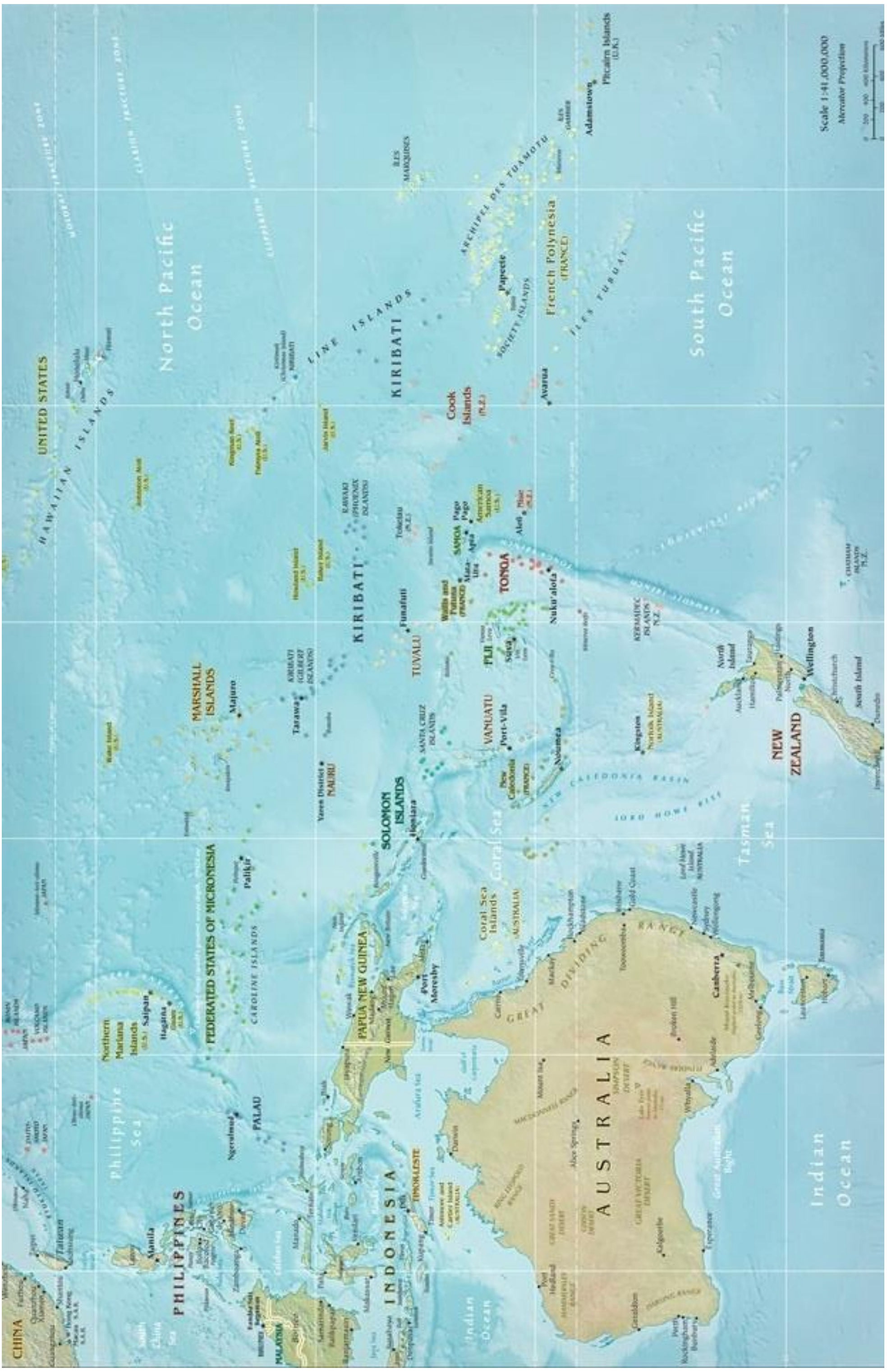
5.3. C1b-M38 without the C1b-M208 Mutation (C1b-M38 xM208).

C1b-M38 xM208 represents a common marker found among the populations of eastern Indonesia and Papua New Guinea (e.g., Mona et al. 2007; Mona et al. 2009; Karafet et al. 2010; Tumonggor et al. 2014). Karmin et al. (2015) suggest that C1b-M38 evolved about 24 thousand years ago. Based on their interpretation of the data, Mona et al. (2007) suggest that this occurred in the northwestern part of New Guinea, and the mutation eventually expanded both to the eastern part of the island (Papua New Guinea) as well as westward to Indonesia. For geneticists, unidentified C1b-M38 mutations that evolved later represent an indigenous component within the genetic tapestry of Island Southeast Asia. For linguists, they are informative markers for deciphering the evolution of the so-called "Papuan" macro-family of languages, a topic that will be discussed in greater detail in Chapter 14.

5.4. C1b-M208 and Austronesian.

The C1b-M208 mutation is a downstream variant of C1b-M38 that evolved about 12 thousand years ago in the highlands of western New Guinea (Delfin et al. 2012; Karmin et al. 2015). C1b-M208 is rarely found in Indonesia (e.g., Mona et al. 2009; Karafet et al. 2010), which suggests a minimal westward expansion of the mutation. However, Mirabel et al. (2012) report that the C1b-M208 mutation exhibits an increasing frequency cline from New Guinea to Rapa Nui. See, also, **Supplementary Tables 6.3** and **6.4**).

Figure 6.2. Island Southeast Asia, Australia, and Oceania. Source: United States Government and Central Intelligence Agency.



The Austronesian language family will be discussed in Chapters 14 and 16. At this point it is important to note that the C1b-M208 mutation is an important marker for linguists because it helps to explain the expansion of Austronesian languages across a vast ten-thousand-kilometer expanse of ocean, from New Guinea to Rapa Nui (Easter Island). Additionally, the C1b-M208 mutation is part of the evidentiary picture that builds the following argument: language continuity can survive population replacement.

Section 6. C1b-M347 and Australian Languages.

Nagle et al. (2016a) present the most comprehensive Y-chromosome study of Aboriginal Australians. According to the study, the C1b-M347 mutation represents about 42 percent of the indigenous Y-chromosome variation within this population. The discovery of the Australian-specific C1b-M347 mutation was initially reported in Hudjashov et al. (2007). Researchers utilized the enhanced resolution of downstream variation within paragroup C-M130 to address a previous study (Redd et al. 2002) that reported Holocene geneflow between India and Australia about 10 thousand years ago. The 2002 study based their findings on a type of genetic marker called short tandem repeats (STR's). Hudjashov et al. (2007) disagreed with the 2002 study and asserted that the Australian aborigines had not experienced any outside geneflow for a period of roughly 45 thousand years, from time that the continent was initially colonized by modern humans until the arrival of Europeans in the late eighteenth century. A similar conclusion was reached by Nagle et al. (2016a).

The observation that aboriginal Australians had not experienced any outside geneflow for a period of roughly 45 thousand years, and the observation that they are the descendants of the human colonization of Australia, presents a salient point for linguists. The Australian language family has roots that extend to the out-of-Africa exodus. Taking this a step further, language must have evolved at least 100 thousand years ago in Africa. This weighs a more plausible assumption, that the out-of-Africa tribe already had language, versus a less plausible alternative explanation, that language evolved independently in several regions of the world.

A complete discussion of indigenous Y-chromosome variation among the Australian aborigines necessitates a discussion of unknown KR-M526 mutations as well as the M-P256 and S-B254 haplogroups. Accordingly, this topic continues in Chapters 13 and 14.

Section 7. C1a-CTS11043 Mutations and the Northern Dispersal Hypothesis.

7.1. Overview of the Archaeological Evidence.

The available archaeological and genetic data suggest that the human tribe split roughly 50 thousand years ago somewhere in the general vicinity of the Black Sea. As detailed above in Section 3 (above), one group migrated southwards and eventually colonized South Asia, Island Southeast Asia, and Australia. This expansion followed what is commonly described in the literature the "southern route" (e.g., Mellars 2006; Stoneking and Delfin 2010; Oppenheimer 2012). However, dispersals along a northern route also occurred. The data suggest that a group of hunter-gatherers migrated northwards from the Caucasus to Eastern Europe or perhaps Central Asia. The group then separated with some migrating to Western Europe, others to East Asia, and others into the Arctic Circle.

The northern dispersal was facilitated by the so-called "mammoth steppe," a unique ecosystem that extended across Eurasia until the beginning of the Holocene (e.g., Ricankova et al. 2014). Modern humans were drawn to this region about 45 to 50 thousand years ago because of the availability of high-quality food. As implied by its name, this vast ecosystem supported a variety of large herbivores that not only included mammoths, but also reindeer, woolly rhinoceroses, wild horses, and bison (e.g., Dolukhanov 2003; Gordon 2003). The archeological record dates the harvesting of large herbivores on the mammoth steppe to at least 45 thousand years ago. The data come from the remains of a woolly mammoth that humans had killed and butchered near Sopochnaya Karga, a meteorological station in Siberia (see Pitulko et al. 2016).

Robust archaeological support for human expansions via the northern is also provided by human remains with the C1-F3393 and C1a-V20 mutations, (see **Supplementary Table 6.5** and in particular, the data for Marine Isotope Stage 3). From an archeological perspective, one salient observation about these individuals is food remains that were identified at the archaeological sites. These data consistently show that large herbivores were on the menu. Taking this a step further, the opportunity to harvest a huge nutritional return for comparatively little time and effort may well explain why these hunter-gatherers expanded away from the Black Sea. As shown by **Supplementary Table 6.5**, important archeological support for this position comes from the Bacho Kiro Cave in Bulgaria (Hublin et al. 2020). The diet of Paleolithic peoples who lived hers included bison and horses. Data also comes from the from the Sunghir archeological site, which is located about 190 km northeast of Moscow. Here, the remains of five males were found. They died between 32 and 34 thousand years ago. Grave artifacts suggest they hunted mammoths (Sikora et al. 2017: S1). Additional support comes from Kostenki 14, an individual discovered along the banks of the Don River in the Voronezh Oblast of Russia. He died about 37 thousand years ago. Animal remains from the Kostenki archaeological site suggest that he ate reindeer and horse (Anikovich et al. 2007). Turning now to Vestonice 16, his remains were discovered at the Dolni Vestonice archaeological site in the Czech Republic. He died about 30 thousand years ago. Animal remains suggest that his diet included mammoth and reindeer (Svoboda et al. 2009). Finally, the Goyet Q116-1 remains were found at an archeological site in Belgium. He died about 35 thousand years ago. Animal remains suggest that his diet may have included horse and reindeer (Stevens et al. 2009).

A recent paper (Bocherens and Drucker 2021) utilized a novel approach to determine the diet of Neanderthals and modern humans who lived in Europe during Marine Isotope Stage 3. The data was obtained from stable isotope analysis of the remains of Neanderthals, modern human, large herbivore, and carnivores that died roughly from 30 to 45 thousand years ago. The analysis indicates that during Marine Isotope Stage 3, the diet of modern humans consisted largely of herbivores that included reindeer, muskox, bison, woolly rhinoceros, horse, and mammoth. Among these herbivores, mammoth provided the largest source of protein (see Figure 13.3 from the study for an informative summary of the results).

7.2. Genetic Support for the "Northern Dispersal" Hypothesis.

It should be note that the C1a-V20 mutation remained part of the European genome until the end of the Neolithic. It has since disappeared from the continent. Among contemporary global populations, C1a-CTS11043 mutations are only observed among the Japanese, and in particular, the C1a-M8 downstream variant. It should be noted that C1a-M8 occupies a phylogenetically equivalent position to C1a-V20 within the Y-chromosome tree.

At this point the reader is directed to **Supplementary Figure 6.1**. As shown by the figure, the C1a-M8 mutation (red border) and C1a-V20 mutation (brown border) diverge from C1a-CTS11043. According to Poznik et al.), both mutations evolved about 48 thousand years ago, a figure that corresponds well with archeological support for dispersals along the northern route.

As shown by **Supplementary Table 6.5**, the C1a-V20 mutation was found in Sunghir, Goyet Q116-1, and Vestonice 16 remains which date to Marine Isotope Stage 3. However, the Kostenki 14 individual, important remains from this period, has the phylogenetically distant C1b-F1370 mutation. This is unexpected because C1b-F1370 is the genetic signature of human expansions via the southern route during Marine Isotope Stage 3 and the Kostenki 14 remains were found near the Black Sea. The most parsimonious explanation is that the Levant defines the geographic point of dispersal for human migrations during Marine Isotope Stage 3. C1a-CTS11043 and C1b-F1370 evolved in this area roughly 50 thousand years ago. Two or three thousand years later, populations with these mutations rapidly expanded from the Black Sea because of climate change.

Section 8. C1a-M8 and Japonic Languages.

The C1a-M8 mutation is observed in about six percent of contemporary Japanese (Sato et al. 2014). As explained in Chapter 4 and the discussion of haplogroup D-M174, the D1b-M55 mutation represents a Paleolithic component of the contemporary genome in Japan. This mutation stands a genetic relic of the human colonization of the Japanese Islands roughly 30 thousand years ago. A similar conclusion can be made for the C1a-M8 mutations. However, the genetic and archaeological data suggest that D1b-M55 arrived in the Japanese Islands via the southern route. C1a-M8, on the other hand, probably arrived in Japan via the northern route.

For linguists, the salient point here is that the prehistoric Jomon culture may have contributed to the evolution of Japonic languages, a discussion that will continue in Chapter 16. The C1a-M8 mutation was also observed in a Mesolithic individual from Laos (see **Supplementary Table 6.5**). This suggests that C1a-CTS1103 mutations occupied a larger section of East Asian genome prior to the onset of the Neolithic.

Section 9. Conclusions for Haplogroup C1-F3393.

Y-chromosome haplogroups are unique segments of human genetic diversity. Data that has accumulated over the last twenty years require a revision of the Y-chromosome phylogeny; namely, the creation of a C1-F3393 haplogroup. The C1-F3393 haplogroup has two main divisions, C1a-CTS11043 and C1b-F1370. C1b-1370 mutation expanded out of the levant during Marine Isotope Stage 3. Today, these mutations represent a Paleolithic component among the populations of South Asia, East Asia, Island Southeast Asia, and Australia, and a Neolithic component in Oceania. For linguists, C1b-F1370 provides especially informative mutations for deciphering the prehistory of Australian, Papuan, and Austronesian languages. C1a-CTS11043 mutations, on the other hand dispersed across Eurasia during Marine Isotope Stage 3 via the northern route. These mutations occupied a significant corner of the Eurasian genome until the Neolithic. Today, C1a-CTS11043 only attains a significant frequency among the Japanese.

Chapter 7: Haplogroup C2-M217.

Section 1. Contemporary Distribution of Haplogroup C2-M217.

Like the C1-F3393 haplogroup, the designation of C2-M217 as a haplogroup also represents a departure from the standard Y-chromosome nomenclature (see Y-Chromosome Commission 2002 and the International Society for Genetic Genealogy).

The reader is now directed to **Supplementary Table 7.1**. As shown by the table, the C2-M217 haplogroup attain a high frequency among the so-called Altaic-speaking populations of Central Asia, East Asia, and Northern Eurasia. Moderate frequencies of the haplogroups are observed among Chinese-speaking populations in East Asia. Low frequencies are observed among Koreans, Japanese, and Native Americans.

Section 2. Evolutionary History of Haplogroup C2-M217.

2.1 Overview.

As shown by **Supplementary Figure 1.1** from the first chapter, haplogroups C1-F3393 and C2-M217 diverge from the C-M130 paragroup. As previously discussed in Chapter 6, this occurred roughly 49 thousand years ago. Nevertheless, the C1-F3393 and C2-M217 haplogroups have evolutionary histories that are quite different. This explains why the contemporary distribution of both mutations is vastly different. Moreover, this also explains why C1-F3393 mutations are important mutations for deciphering the prehistory of Japonic, Papuan, Austronesian, and Australian languages, and why C2-M217 mutations help to decipher the prehistory Altaic and Native American languages.

As detailed in Chapter 6, haplogroup C1-F3393 represents the genetic relic of human migrations during Marine Isotope Stage 3, roughly 50 thousand years ago. The dispersal of C1-F3390 mutations followed two different migratory trajectories, a southern route, and a northern route. These expansions signal the initial human settlement of South Asia, Island Southeast Asia, Australia, Japan, and Europe. However, the geographic expansion of haplogroup C2-M217 during Marine Isotope Stage 3 was more restricted. The available data suggest that hunter-gatherers with the mutation migrated from the Black Sea into East Asia about 50 thousand years ago. Somewhere in the general vicinity of northern Mongolia populations with this mutation "nested" until Pleistocene/Holocene transition that began roughly 20 thousand years ago.

2.2. Genetic Perspectives.

As discussed previously in Chapter 4, the DR-M168 mutation represents the exodus of modern humans from Africa into the Levant roughly 100 thousand years ago. In this region, roughly 70 thousand years ago, D-M174, E-M96, and C-M130 diverged from DR-M168. As noted above, haplogroups C1-F3393 and C2-M217 diverged from the C-M130 paragroup about 49 thousand years ago.

The reader is now invited to review **Supplementary Figure 7.1**. As shown by the figure, the C2b-L1373 and C2c-F1067 mutations form the two main divisions of the haplogroup C2-M217 downstream phylogeny. According to Wu et al. (2020), C2b-L1373 and C2c-F1067 diverged from C2-M217 about 34 thousand years ago. These data support the archaeological discussion (below) with the idea that C2-M217 and Paleolithic hunter-gatherers thrived and survived in the general vicinity of northern Mongolia. Taking this step further, the timing of the C2b-L1373 and C2c-F1067 split defines northern Mongolia and Lake Baikal geographic location where the initial diversification of C2-M217 occurred. Additionally, the ancient C2-M217 data, although limited, offer support for this position. This stems from the distribution pattern of ancient C2-M217 mutations in Eurasia (see **Supplementary Table 7.2**). Finally, this position is supported by the contemporary distribution of haplogroup C2-M217 in Eurasia (see **Supplementary Table 7.1**).

Figure 7.1. Mongolia and Lake Baikal.

2.3. Climatological and Archaeological Perspectives.

Archaeological and climate support for Marine Isotope Stage 3 dispersals via the northern route was previously detailed in Chapter 6. These data posit a human expansion from the Black Sea roughly 50 thousand years ago. Somewhere in Eastern Europe or perhaps Central Asia, the human tribe separated. Some migrated to Western Europe, others to East Asia, and others in the direction of the Arctic Circle. Hunter-gatherers with the C2-M217 mutation appear to have settled in the general vicinity of northern Mongolia and Lake Baikal in Russia roughly 45 thousand years ago. Archaeological support for this conclusion stems from a recent paper (Zwyns et al. 2019) which details numerous archeological sites within this region The paper also provides dating results which suggest that hunter-gatherers arrived in this region roughly 45 thousand years ago.

The archeological complex discussed by Zwyns et al. (2019) appears to have collapsed at the onset of Marine Isotope Stage 2 and the Last Glacial Maximum, roughly 27 thousand years ago. Data from Rybin et al. (2016) suggest that at this point in the climate record may have also pushed hunter-gatherers and C2-M17 mutations southwards into northern China. Additionally, it should be noted that deglaciation at the onset the Holocene may have pushed C2-M217 mutations into the Russian Far East and the Amur River Basin. This conclusion is based on archaeological data provided by Buvit and Terry (2011).

Section 3. Altaic Languages.

3.1. Overview.

Turkic, Tungusic, and Mongolic are currently recognized by Ethnologue and Glottolog as language families. However, as summarized by Campbell and Poser (2008: 235-241), the so-called "Altaic hypothesis" has circulated among linguists for at least 150 years. This hypothesis posits that the Turkic, Tungusic, and Mongolic language families evolved from a common ancestral language. According to both researchers, proponents of the Altaic classification cite common features that include vowel harmony, SOV word order, agglutinative morphology, the absence of the verb "to have," and the absence of grammatical gender. Arguments against the Altaic hypothesis include a lack of cognates for basic vocabulary, such as cognates for body parts. As such, similarities found in Turkic, Tungusic, and Mongolic could be potential relics of intense language contact over a prolonged period, and with that, Altaic becomes an East Asian *Sprachbund*.

3.2. Linguistic Perspectives and Turkic.

Glottolog (4.4) lists 44 languages within the Turkic language family. Moreover, Turkic-speaking populations are spread over a wide geographical expanse that extends from Eastern Europe to East Asia. Examples of Turkic languages include the following: Gagauz in Eastern Europe; Turkish in Southwest Asia; Azerbaijani in the Caucasus; Kazakh, Kyrgyz, Turkmen, and Uzbek in Central Asia; Änyu in East Asia; Yakut and Dolgan in Siberia. Interestingly, the where and when of Turkic languages still remains a mystery (e.g., Kornfilt 2009). The earliest attestations are dated to the eighth or ninth century and consist of Orkun Inscriptions from Mongolia and Old Uyghur manuscripts from in Xinjiang, China. As such, the putative homeland of Turkic languages might lie somewhere in East Asia.

3.3. Linguistic Perspectives and Mongolic.

Ethnologue (2017) classifies thirteen languages within the Mongolic language family. Twelve of the languages are spoken either in China, Russia, or Mongolia. The other Mongolic language, Mogholi, is found in Afghanistan. The earliest attestation of Mongolic languages is the so-called "Para-Mongolic" Khitan scripts dating to about the tenth century (Kane 1989: 11-37; Janhunen 2003a: 394-396), which were prepared during the Liao Dynasty. Pre-Classical Mongolic texts later emerged during the reign of Genghis Khan in the thirteenth century (Janhunen 2003b: 32-33).

3.4. Linguistic Perspectives and Tungusic.

According to *Ethnologue* (2017), the Tungusic language family consists of eleven languages spoken by around 55 thousand speakers either in northeastern China or eastern Siberia. Determining the putative homeland of Tungusic is complicated by the contemporary and historical distribution of this language family. Tungusic-speaking populations include small populations in Siberia, such as the Even and Evenki, whose survival strategy once included the domestication of reindeer. In contrast, another Tungusic language, Manchu, stands as a former linguistic heavyweight, a relic of the Qing Dynasty of China. When the dynasty collapsed in 1912, the Manchu language rapidly became moribund.

The earliest attestation of Tungusic stems from texts that appeared in the twelfth century during the Chinese Jin dynasty. These texts were written in the Jurchen language using characters borrowed from Khitan (a Mongolic language) and Chinese (Kane 1989:1-10).

3.5. Genetics and Altaic.

Turning now to linguistically informative mutations, **Supplementary Figure 7.1** highlights (with a red border) significant C2-M217 mutations that are especially prevalent among Altaic-speaking populations: C2b-M48, C2b-F1918, C2c-CTS2657, and C2c-F8465. Frequency data for C2b-M48, C2b-F1918, and C2c-CTS2657 are found in **Supplementary Tables 7.3**, **7.4**, and **7.5**). These data suggest significant admixture among Turkic, Tungusic and Mongolic populations over a prolonged period of time. Taking this a step further, the data provide strong support that defines the evolution of Altaic languages according to a *Sprachbund* model. This follows the heavy frequencies of the mutations observed among Altaic-speaking populations versus their absence among non-Altaic populations. Notable exceptions to this pattern are the frequency of C2b-M48 among the Yukaghirs, and C2b-F1918 among the Hazara. Yukaghir is a language family, and Hazara is an Indo-European language.

3.6. Genetics and Tungusic.

A recent Y-chromosome study (Liu et al. 2020) explored the phylogeny of the C2b-M48 mutation. According to the researchers, C2b-F7171 mutations represent C2b-48 variation among Turkic and Mongolic-speaking populations. C2b-F5484 represents C2b-M48 variation among Tungusic-speaking populations. Based on their analysis of the genetic data, the study further suggests that Tungusic languages evolved in Amur River Basin of the Russian Far East and expanded out of this region about three thousand years ago.

Haplogroup N-M231 is also frequently observed among Tungusic-speaking populations. Accordingly, this discussion continues in Chapter 15.

3.7. Genetics and Mongolic.

A study from 2017 (Huang et al.) examined C2-M217 mutations downstream from C2c-CTS2657. According to the study, C2c-F8465, a downstream variant of C2c-CTS2657, represents the genetic signature of Mongolic languages. The study further reports that this mutation evolved roughly 4,000 years ago in northeastern Asia.

3.8. Genetics and Turkic.

According to the Y-chromosome perspective, language contact theory provides an especially robust model for explaining the expansion of Turkic languages across Eurasia. For example, around 30 percent of Turkic speakers live in Turkey (Kornfilt 2009). Interestingly, a Y-chromosome study from 2004 (Cinnioglu et al.) reports that the haplogroup C2-M217 attains a frequency of less than 1 percent of Turkish males. This suggests that language shift occurred in Anatolia without significant admixture with Turkic-speakers from Central Asia or Northern Eurasia. Such a conclusion agrees with the historical record and the demise of the Byzantine Empire. This underscores the following: language expansion can occur without a population expansion. Another example is the Yakuts, a Turkic-speaking population of Siberia. Their reliance on reindeer herding and the high frequency of haplogroup N-M231 mutations suggest that they initially spoke a Uralic language (e.g., Pakendorf et al. 2006). Another example is the Gagauz, a Turkic-speaking population in East Europe. Their Y-chromosome profile is similar to other populations in the Balkans region (see Varzari et al. 2009).

3.9. Climate and Archaeological Perspectives.

Several studies have explored the downstream phylogeny of informative Y-chromosome mutations among Altaic-speaking populations. Dating estimates from these studies suggest that the evolution of Altaic languages is strongly linked with the evolution of agriculture on the East Eurasian steppes. See Huang et al. (2017) and the discussion of C2c-CTS2657; Wei et al. 2018(b) and the discussion of C2b-F1918; Liu et al. (2020) and the discussion of C2b-M48. It appears as though agriculture fueled greater reproductive success, which fueled rapid diversification of Altaic-specific C2-M217 lineages. A recent paleogenomic study (Cui et al. 2020) builds on this idea and suggests that the evolution of Altaic languages is linked to cultivation of Millet that evolved in northern China roughly 8,000 years ago. However, the archaeological record fails to support the cultivation of millet as a significant source of food for this region (e.g., Stevens and Fuller 2017). Rather, hunting and gathering provided most of the calories until the domestication of the horse.

The agricultural transition on the East Eurasian steppes was a gradual process that involved three major developments. The first development was the domestication of the horse, which initially occurred about 5,500 years ago in north-central Kazakhstan (Frachetti 2012). It should be noted that during the Pleistocene, horses were one of several large mammals hunted by the cultures of the mammoth steppe. Thus, the domestication of the horse could be seen as an effort to ensure its continued availability as a source of food. The second development (see, once again, Frachetti 2012) occurred around 4,500 years ago when cattle, goats, and sheep became part of the subsistence strategy of this ecoregion. The final development occurred 3,200 years ago (see Taylor et al. 2020). At this point the cultures of the region had perfected horseback riding. This development allowed pastoralists to move their herd animals over a much greater geographical expanse in search of water and forage. As a result, pastoralism became a more efficient subsistence strategy.

Herd animals provided the Neolithic and later cultures of the East Eurasian steppes with meat, wool, and hide. Additionally, these animals provided dairy products such as milk, curds, and *airag*, a fermented alcoholic beverage produced from horse milk. A recent study (Wilken et al. 2020) provides useful discussion of the evolution of dairy pastoralism on the East Eurasian steppes. The study generates additional archaeological evidence that times the completion of the agricultural transition in the region to roughly 3,000 years ago.

3.10. Conclusions for Section 3.

The use of "Altaic" in the above discussion should not be interpreted as a desire to resurrect the former Altaic language family classification. This research guide generally defers to consensus among the linguists with regard to classification and there appears to be a lack of enthusiasm for support for the idea. Rather, the use of Altaic facilities an elegant presentation of the data that ultimately support an East Eurasian *Sprachbund*. In summary, a synthesis of the linguistic, archaeological, genetic, and climatological perspectives strongly suggests that Turkic, Tungusic, and Mongolic evolved on the East Eurasian steppes. Moreover, these data strongly suggest a period of intense contact among speakers of these languages in the "general vicinity" of Mongolia.

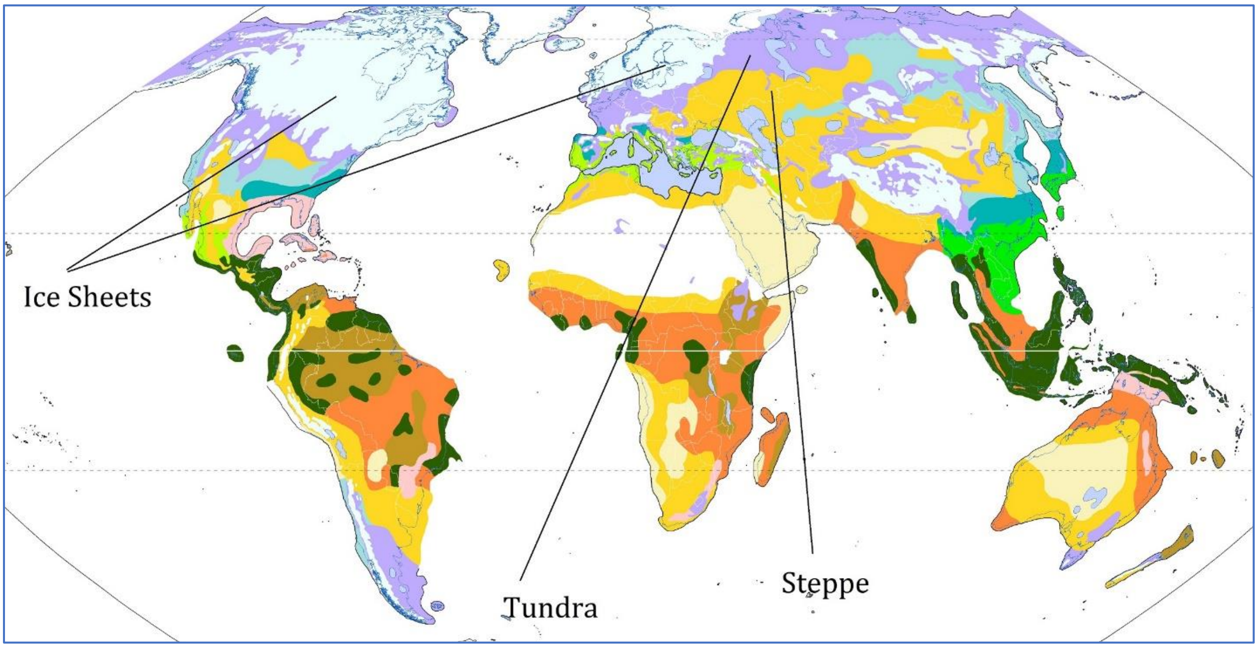
Section 4. C2-M217 and the Non-Altaic Languages of East Asia.

During Marine Isotope Stage 3 the ice glaciers retreated. This facilitated a rapid expansion of the human tribe across Eurasia. However, the weather deteriorated at the onset of Marine Isotope Stage 2, roughly 29 thousand years ago. The ice glaciers eventually reached their maximum southern extent in the northern hemisphere about 27 thousand years ago, a point that roughly equates and fifty degrees north in much of Central and East Asia. The advance of ice glaciers curtailed human migration and populations expansions. Rather, human populations settled in several refugia across the Eurasian landmass where they waited for better weather.

The glaciers began to recede roughly 19 thousand years ago. Some populations, such those as in present-day Japan and Australia, remained in-place. Populations in other refugia expanded. For geneticists, the isolation of populations during the Last Glacial Maximum, and their subsequent post-glacial expansion at the end of the Pleistocene, or alternatively, their continued isolation into the Holocene, represents a partial explanation for contemporary global genetic diversity. For the linguists, this provides a partial explanation for global linguistic diversity.

An interesting study from 2016 (Gavashelishvili and Tarkhnishvili) used computer simulation to identify the refugia where human survived during the Last Glacial Maximum. Additionally, they identified the human Y-chromosome haplogroups that expanded from these refugia with the onset of the Holocene. Their model was constructed utilizing a synthesis of climate, terrain, and hydrographic data, as well as data from fossilized pollen and plant remains. Their analysis of the data places many of the Eurasian refugia within the tundra and steppe belt directly south of the glacial ice sheets. As such, adaptation to cold weather (cold adaptation) enabled the human tribe to thrive and survive during this period.

Figure 7.2. Ice Sheets, Tundra, and Steppe 20 thousand Years Ago. Source: Wikipedia and Fährtenleser.



Dolukhanov (2003) suggest that cold adaptation succeeded on the East Eurasian steppes because arid conditions produced, at worst, a thin layer of snow and ice on the ground. Consequently, the steppes continued to provide an ideal habitat for a variety of large herbivores such as mammoths, woolly rhinoceros, wild horses, and bison. Even during the winter months these animals could easily forage as they simply had to scrape away a thin layer of snow to access the grass underneath. This explains the continued success of Ice Age hunter-gatherers during Marine Isotope Stage 2. The large herbivores continued to thrive, and the hunters continued to feast on an abundant source of protein that could be harvested at a comparatively small expenditure of energy. Of course, the mammoth hunter tradition eventually ended at the onset of the Holocene, both in Eurasia and in North America, and once again, the human tribe was forced to adapt.

The above discussion of the Last Glacial Maximum, and the Holocene transition that followed, potentially explains the moderate frequency of C2-M217 mutations observed among the non-Altaic populations of East Asia. These mutations represent a genetic relic of Ice Age refugia in East Asia. Since this topic requires a discussion of haplogroup N-M231, more details will be provided in Chapter 15. In the meantime, it should be noted that C2-M217 mutations among Altaic populations are different than those of Chinese, Koreans, Japanese, and Vietnamese. As shown by **Supplementary Figure 7.1**, the currently available data suggest that Altaic C2-M217 mutations essentially occupy unique regions of the Y-chromosome "map."

Section 5. C2-M217 and the Transeurasian Hypothesis.

As noted previously, the idea of an Altaic language family posits the evolution of Tungusic, Mongolic and Turkic from a common ancestral proto language. It is difficult resolve language classification controversies like this with non-linguistic data. Traditionally, classification reflect requires consensus among the linguists, and many linguists do not endorse the Altaic classification based on the lack of linguistic evidence. Nevertheless, as noted previously, a synthesis of different data sources supports the concept of Altaic *Sprachbund*. Many linguists are probably comfortable with this idea.

At the Max Planck Institute for the Science of Human History, a research group is currently exploring the Transeurasian hypothesis. This project proposes a common Transeurasian proto language that unites the origins of Altaic languages with that of Japonic and Koreanic (see Robbeets 2017a for an overview). From a Y-chromosome perspective, the C2-M217 data fail to support their position. Additionally, the C2-M217 data fails to build an alternate language contact model that would unite Japonic and Koreanic with Altaic.

Evaluation of the Transeurasian hypothesis also requires analysis of haplogroup N-M231 mutations among Koreans and Japanese. Accordingly, this discussion continues in Chapter 16.

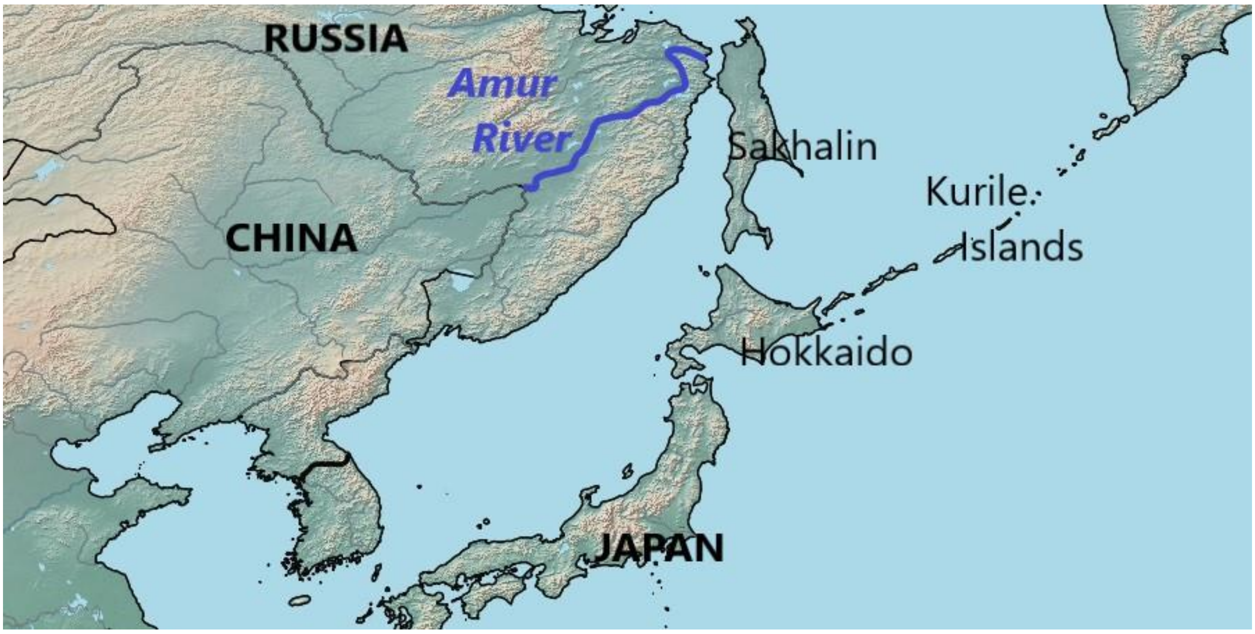
Section 6. C2-M217, Ainu, Japanese, and the Amur River Basin.

The C2-M217 haplogroup has been detected in Mesolithic remains from the Devil's Gate Cave, which is in the Russian Far East (see **Supplementary Table 7.2**). The haplogroup is also observed in about 6 percent of the Japanese (Sato et al. 2014). Likewise, the mutation is observed among the Ainu people on the Japanese island of Hokkaido. According to Tajima et al. (2004b), the frequency of C2-M217 among this population is about 17 percent. However, this may not reflect the actual frequency because of the small sample size of 16 Ainu.

The C2-M217 data for the Russian Far East and Hokkaido may support prehistoric geneflow between both locations. Archaeological support for this conclusion comes from the distribution of so-called micro-blade tools. A 2015 report (Yi et al.) provides a useful discussion of this tool making tradition. According to the researchers, micro-blade tools suddenly appeared in northern China, the Korean Peninsula, the Russian Far East, and northern Japan around the onset of the Last Glacial Maximum. According to Yi et al. (2015), this technology was ideally suited for the needs of the cold-adapted hunter-gatherers of this period. Prior technology for making a knife utilized a large stone blade that was knapped from a larger stone. Mico-blades, on the other hand, are fashioned from a series of small sharp wedges that inserted into a wood or bone shaft. These knives have several advantages: lighter weight; resistance to shattering in the extreme cold; and the ability to utilize more readily available stone of lesser quality.

Takahura (2012) suggest that the micro-blade tradition spread from the East Asian mainland to Hokkaido by around 20 thousand years ago. In a report from 2020, Takahura provides an updated discussion of micro-blade tools found at archaeological sites on Hokkaido. This tradition in northern Japan appears to have ended around the beginning of the Holocene, roughly 11 thousand years ago. According to the 2020 report, climate change may have eliminated the need for micro-blade tools.

Figure 7.3. Amur River, Sakhalin, and Hokkaido.



Prehistoric cultural exchange between the Russian Far East and Hokkaido may have also involved the obsidian trade. Obsidian, or volcanic glass, was utilized by prehistoric cultures worldwide as a raw material for making blades. The cutting ability of these blades apparently rivals that of a modern surgical scalpel. Glascock et al. (2011) presented geo-chemical analysis of obsidian artifacts found at archeological sites in the Amur River Valley. According to the analysis, the obsidian was sourced from Hokkaido. Archeological data from the study further suggests that this material was transported to the Amur River Basin over a distance of a thousand kilometers via the Sakhalin Islands. According to the archaeological data, this occurred between 7,200 and 8,600 years ago.

It should be emphasized that exchange networks can produce geneflow and language expansion, but this is certainly not the only possible outcome. Nevertheless, C2-M217s mutations in Japan and the Russian Far East, and the archeological evidence of prehistoric cultural contact between both regions, suggest that the prehistoric Jomon culture of Japan was less isolated than previously assumed by some researchers. This observation serves a linguistic purpose that becomes clearer in Chapter 16 and the discussion of haplogroup O-M175.

Section 7. C2-M217, Native Americans, and Koryaks.

Among the indigenous populations of North America, haplogroup Q-M242 carries about 93 percent of the indigenous genetic component, whereas C2-M217 represents the remaining 7 percent (e.g., Zegura et al. 2004). Moreover, Q-M242 represents almost all of the indigenous Y-chromosome variation in South America (Geppert et al 2011; Roewer et al. 2013; Jota et al. 2016). C2-M217, on the other hand, rarely surfaces among the indigenous populations of the continent. Moreover, Pinotti et al. (2019) report that indigenous South Americans have a C2-M217 variant that is evolutionary distant from that found among the indigenous peoples of North America. These researchers further report that the unique South American C2-M217 variant and the unique North American C2-M217 variant diverged from a common ancestor roughly 22 thousand years ago.

Based on the available data, both ancient and contemporary, C2-M217 mutations are especially helpful for deciphering the prehistory of the Eskimo-Aleut, Eyak-Athabaskan, and Chukotka-Kamchatkan language families. This follows the idea that the evolution of these languages was influenced by prehistoric geneflow across the Bering Sea. Since this topic also required analysis of data from the Q-M242 haplogroup, the story of Koryaks and Native Alaskans continues in Chapter 17. In the meantime, the reader is invited to examine **Supplementary Figure 7.1**. The C2b-P39 mutation found among Native Americans is highlighted by a blue border. As shown by the figure, the C2b-FGC28881.2 mutation is a phylogenetic sister clade of Cb2-P39. C2b-FGC28881.2 is found among the Koryaks (see Wei et al. 2017b). Among the Paleo-Siberian peoples of Asia, Koryaks have traditionally lived along the Bering Sea near the Kamchatka Peninsula. They speak a language belonging to the Chukotko-Kamchatkan language family. Moreover, they have traditionally employed a hunter-gather subsistence strategy that included the harvesting of whales, a cultural adaptation that links them with the prehistory Native Alaskans (see Chapter 17).

Surprisingly, the oldest remains with the C2-M217 mutation come not from Asia but rather from a Native American who died about 10 thousand years ago in Brazil (see **Supplementary Table 7.2** for more details). It should be stressed that this discovery does not support the evolution of C2-M217 in South America. Conclusions such as this must be drawn from several different data sources and the data clearly point to the evolution of C2-M217 in East Asia. Its presence in the remains from Brazil is best explained by Upper Paleolithic migrations into the Americas, a topic that is explored in Chapter 17. Moreover, the data from Brazil illustrate that obtaining well-preserved DNA from human remains is difficult. In the end, finding sufficiently preserved DNA for sequencing is a matter of luck. Perhaps the future will bring additional ancient C2-M217 data for Eurasia.

Section 8. Conclusions for Haplogroup C2-M217.

The evolutionary history of the C1-F3393 and C2-M217 mutations is so vastly different that they should be defined as haplogroups. C2-M217 stands as a genetic relic of the human colonization of East Asia via the northern route. Diversification of this mutation is explained by cold adaptation, the expansion and retreat of ice glaciers, and the Neolithic transition on the East Eurasian steppes. Moreover, the regularity and heavy frequency of C2-M217 mutations among contemporary Altaic population is striking. This observation, along with linguistic and archaeological perspectives, invites linguists to seriously consider the concept of an Altaic *Sprachbund*.

Chapter 8: Haplogroup G-M201.

Section 1. The Contemporary Distribution of Haplogroup G-M201.

The G-M201 haplogroup attains its highest frequency in the Caucasus and Central Asia (see **Supplementary Table 8.1**). Turning now to its internal phylogeny, G-M201 has two main branches, G1-M285 and G2-P287 (**Supplementary Figure 8.1**). As shown by **Supplementary Table 8.2**, the contemporary distribution of G1-M285 is rather limited and is confined almost exclusively to populations in Asia. **Supplementary Table 8.3**, on the other hand, indicates that the contemporary distribution of G2-P287 is much broader, having a range that extends from Western Europe to Central Asia.

Section 2. The Evolution and Expansion of the G-M201 Haplogroup.

At this point the reader is directed to **Supplementary Figure 1.1** from the first chapter. The DR-M168 paragroup represents the ancestral mutation of Y-chromosome haplogroups that evolved outside of Africa. These haplogroups include D-M174 and E-M96 as well as the C-M130 paragroup. The sister clade of C-M130, the FR-M89 mutation, eventually evolved into G-M201 and HR-M578 around 50 thousand years ago (Poznik et al. 2016). This probably occurred directly south of the Caspian Sea (Rootsi et al. 2012).

D-M174, E-M96, C1-F3393, and C2-M217 expanded out of Southwest Asia during the Paleolithic, about 50 thousand years go. G-M201, on the other hand, began to expand out of this region much later, during the Neolithic, roughly 10 to 12 thousand years ago, after agriculture had evolved in Southwest Asia. This observation is supported by analysis of contemporary Y-chromosome data (see Rootsi et al. 2012). As shown by **Supplementary Table 8.1**, the G-M201 haplogroup exhibits a decreasing frequency cline from the Caucasus, both westwards and eastwards. Additionally, ancient G-M201 data first appears in Neolithic remains with the oldest in Southwest Asia and the most recent in Iberia (see **Supplementary Table 8.4**).

Section 3. The Importance of Southwest Asian Neolithic for Linguists.

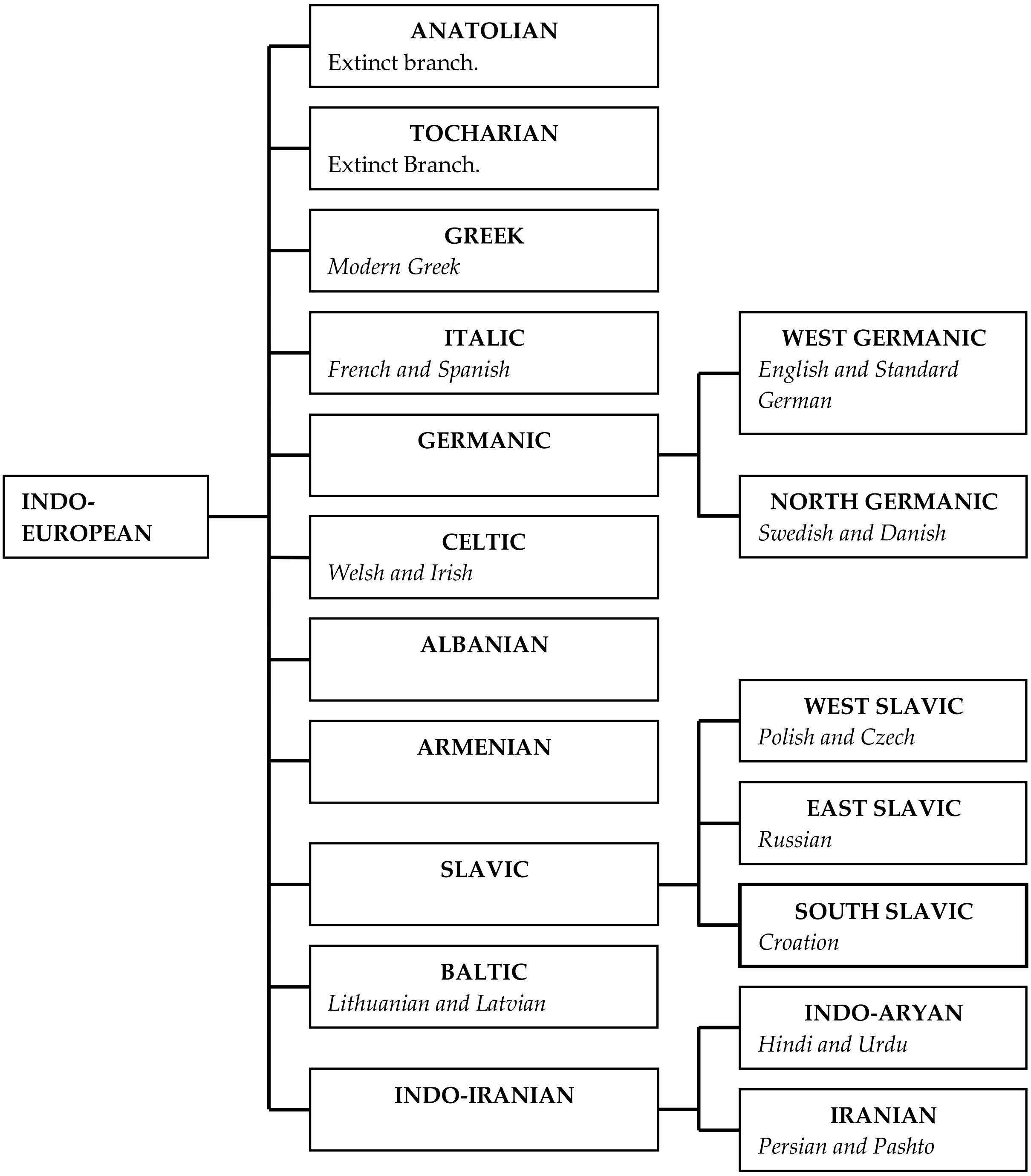
The evolution of agriculture in Southwest Asia and the so-called "Fertile Crescent" was previously introduced in Chapter 5. About 14 thousand years ago people initially harvested wild cereals. This led to a series of innovations that included the development of pottery, the genetic modification of cereals and legumes for cultivation, and the domestication of goats and sheep. As detailed in Chapter 5, the Southwest Asian agriculture package, E-M96 mutations, and Afro-Asiatic languages co-expanded into North and East Africa roughly 6,400 thousand years ago. Turning now to Europe, South Asia, and Central Asia, the Neolithic transformation in all these regions also resulted from an expansion of agriculture from Southwest Asia. However, unlike North Africa, the archeological, genetic, and linguistic relics of the Neolithic transformation in these regions involved a co-expansion of farmers, haplogroup G-M201 mutations, and Indo-European languages.

Section 4. Haplogroup G-M201, the Archaeology Record, and Indo-European.

The Indo-European language family is indeed a linguistic heavyweight with roughly 3.29 billion speakers (Ethnologue). It is difficult to determine the number of Indo-European languages.

Numbers provided by *Ethnologue* (2021) and *Glottlog* (Version 4.4), claim 446 and 583 Indo-European languages respectively. These figures seem, however, overinflated as both catalogs tend to elevate dialects (e.g., Scots English) to the language level. Turning now to the branches of Indo-European, Anatolian and Tocharian are extinct, whereas Greek, Italic, Celtic, Germanic, Albanian, Armenian, Slavic, Baltic, and Indo-Iranian form the extant branches. The reader is directed to Figure 8.1 (below).

Figure 8.1. Indo-European: Branches, Selected Sub-Branches, and Representive Languages.



Most linguists would agree that prehistoric expansions explain the initial dissemination of Indo-European languages into Europe, South Asia, and Central Asia. Nevertheless, deciphering the putative homeland of Indo-European has become a controversial topic among researchers. Most linguists endorse the *steppe nomad hypothesis*. This model links the spread of Indo-European with a Bronze Age expansion of pastoralists from Eastern Europe or Central Asia roughly four thousand years ago (e.g., Gimbutas 1997; Anthony 2007; Anthony and Ringe 2015). Advocates of this hypothesis base their conclusions largely on linguistic reconstructions which, in their opinion, allow researchers to reconstruct a proto-Indo-European culture. However, the archeological record fails to support this model. Rather, the archeological perspective supports a co-expansion of early agriculture and Indo-European languages from the Fertile Crescent during the Neolithic, roughly eight thousand years ago (see Renfrew 1987 and 1989).

The *early farming dispersal hypothesis* was proposed by the archeologist Peter Bellwood in his 2005 monograph *First Farmers: The Origins of Agricultural Societies*. One idea that surfaced in the monograph is that the expansion of early agriculture provides a highly persuasive explanation of how Indo-European spread over a vast geographical expanse, from western Europe to eastern India. Early agriculture links Europe and South Asia with a common cultural transformation that occurred at almost exactly the same time: the expansion of the Southwest Asian agricultural package beginning roughly 8,500 years ago. Moreover, Bellwood's analysis of the archeological data clearly demonstrates that this model of Indo-European origins is not an isolated phenomenon. Rather, Indo-European is one of several different language families throughout the world that co-expanded with the spread of early agriculture.

A detailed previously in Chapter 1, triangulated Y-chromosome-based modeling represents a methodological solution for deciphering the prehistory of language with genetic tools. Application of this methodology supports the *early farming dispersal hypothesis*. In other words, Indo-European is simply not an exception to the rule. From a triangulated Y-chromosome-based perspective, other language-farming expansions include Arawak (Chapter 17), Niger-Congo (Chapters 3 and 5), Afro-Asiatic (Chapter 5 and 11), Dravidian (Chapters 8 and 11), Sino-Tibetan (Chapters 4 and 16), Austro-Asiatic (Chapter 16), Trans-New Guinea (Chapter 14), Uralic (Chapter 15), and Austronesian (Chapters 6, 14, and 16).

From a triangulated Y-chromosome-based perspective, the co-expansion of early agriculture and Indo-European languages are supported by the contemporary distribution of haplogroup G-M201 mutations (see, also, **Supplementary Table 8.1**). In other words, the contemporary distribution of haplogroup G-M201 matches the geographic range of the initial Indo-European language family expansion. Additional genetic support also comes from G-M201 mutations found in Neolithic remains at archeological sites in Europe and the Middle East (see **Supplementary Tables 8.4** and **8.9**).

From an archeological perspective (Bellwood 2005: 67-84), the expansion of agriculture from Southwest Asia to Europe follows the expansion of two different cultural traditions, the Cardial Ware culture, and the Linear Pottery culture (see Figure 8.2 below). The Cardial Ware expansion began about 10 thousand years when farmers from Anatolia settled on the island of Cyprus. From this location farmers later migrated to Crete, Corsica, and Sardinia. Additionally, the Cardial Ware tradition also expanded along the southern Mediterranean coast of mainland Europe, from Anatolia to Portugal. The Linear Pottery expansion began about 8,500 years ago with an expansion from Anatolia into the Balkans. Then, by around 7,500 thousand years ago, the Linear Pottery culture expansion finally terminated at the coastal plain of northern Germany and in the Low Countries. The Neolithic transition in the British Isles occurred about thousand years later, and Scandinavia required an even longer period of time.

Returning now to contemporary genetic data, several G-M201 mutations have been identified as especially strong genetic relics of the Neolithic transformation in Europe: G2a-L91, G2a-L497, G2a-M406, and G2a-M527 (e.g., Keller et al. 2012; Rootsi et al. 2012; Voskarides et al. 2016; Di Cristofaro et al. 2018). See, also, **Supplementary Tables 8.5 to 8.8**. Moreover, ancient haplogroup G-M201 data help to link these mutations with the Neolithic transformation on the continent (see **Supplementary Tables 8.4 and 8.9**).

Figure 8.2. European Neolithic. Source: Wikipedia.

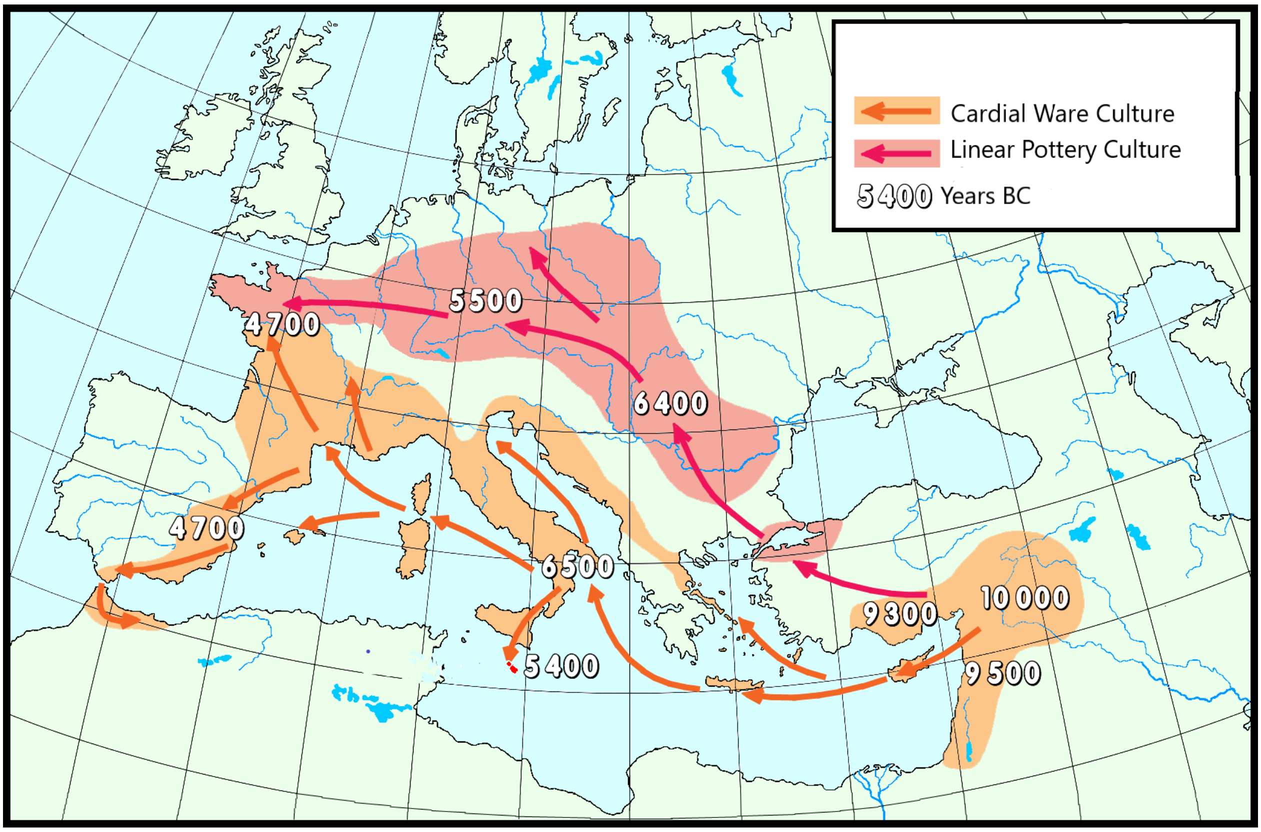


Figure 8.2. European Neolithic. Source: Wikipedia.

Interestingly, a review of the ancient haplogroup G-M201 data has identified samples that came from individuals who met a violent death. The data include Ötzi the Iceman, an individual who died at the end of the European Neolithic. He was murdered in the Italian Alps and shortly thereafter, his body was buried underneath snow and ice. Five thousand years later, in 1991, his remains were discovered by hikers. The cold and altitude left behind a well-preserved corpse that was later identified as having the G2a-L91 mutation (see Keller et al. 2012). Ancient G-M201 remains also include those from an archeological site at Halberstadt-Sontagsfeld in Germany. This site provides evidence of warfare among the farmers of the Linear Pottery tradition (see Meyer et al. 2018). Additional evidence of warfare from this period comes the Schöneck-Kilianstädten site in Germany (see Meyer et al. 2015), as well as Talheim in Germany, and Asparn-Schletz in Austria (see Golitko and Keeley 2007).

Evidence of violence among farmers during the European Neolithic underscores a salient point that linguists should consider. The steppe nomad model of Indo-European languages evolved from the work of Marija Gimbutas, a Lithuanian archaeologist, who proposed the Kurgan conquest model of Indo-European origins in a series of articles she wrote over a forty-year period, ultimately compiled in *The Kurgan Culture and the Indo-Europeanization of Europe: Selected Articles from 1952-1993*. Her theory is often cited in linguistic texts as offering a plausible explanation of how Proto-Indo-European language spread throughout Europe. Trask (1996: 360), for example, writes that while he does not find the Kurgan theory totally persuasive, "it is still the best solution we have and it refuses to go away."

Gimbutas wrote her final article about the Kurgan conquest in 1993, which was published in 1997. This article, "The Fall and Transformation of Old Europe: Recapitulation 1993," reports that the Kurgan culture emerged somewhere in the Volga basin between 5000 and 4500 BC. An identifying trademark of this culture is a unique mortuary practice; they buried their dead in pits, which were then covered with a mound of dirt. In her final article, Gimbutas maintains (1997: 354) that the Kurgans rode horses and raised herd animals within a patriarchal society. Around 4500 BC the Kurgans became more aggressive and began migrating to the west. In the area to the west, what Gimbutas often calls "Old Europe," lived a "Goddess worshiping" culture, whose focus was "the perpetual functioning of the cycle of life, death and regeneration embodied by a central feminine force." (351). Gimbutas asserts (358) that this culture could not resist the Kurgan invasion of warriors from the east who rode horses and who were better armed. During the conquest of Old Europe, the Kurgans imposed their language, Proto-Indo-European, upon the indigenous Europeans (364).

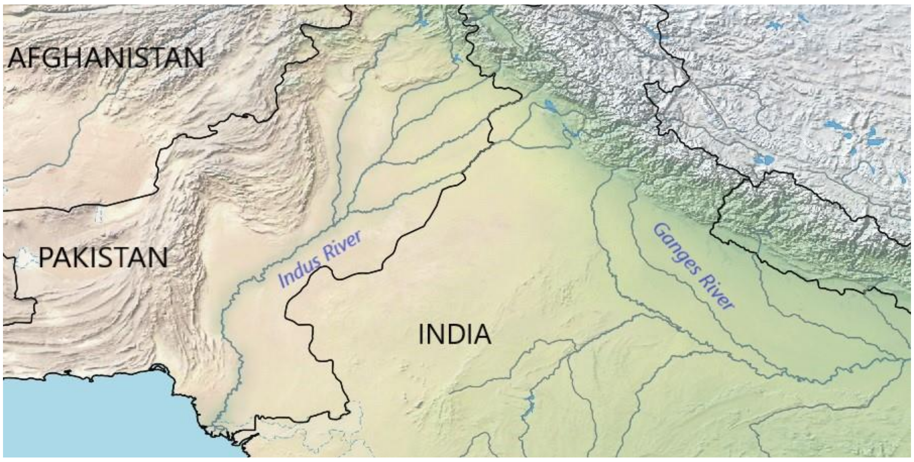
Contrary to what is asserted by asserted by Gimbutas, the archeological record fails support the idea of peaceful Neolithic farmers who were especially vulnerable to attack by aggressive Kurgans. Her work is akin to the concept of the "noble savage" rather than an empirical solution to the Indo-European question.

Section 5. The Indo-Iranian Languages of South Asia.

The *early farming dispersal hypothesis* (Bellwood 2005: 86-95) suggests that the arrival of Indo-Iranian languages in South Asia follows the spread of agriculture from Fertile Crescent during the Neolithic. From an archaeobotanical perspective, the Neolithic in South Asia saw the adoption of Southwest Asian crops such as wheat, barley, lentils, chickpeas, flax, and linseed (e.g., Fuller 2006: 20). Additionally, the Neolithic transition in India also saw the adoption of crops from Africa, such as sorghum and cowpeas (e.g., Crowther et al. 2017), as well as rice from East Asia (see Chapter 16).

The Southwest Asian agricultural package appears to have spread to South Asia via the Iranian Plateau. According to the archaeological record, by around 9,000 years ago numerous farming settlements appeared at Mehrgarh, a Neolithic site located in the Balochistan region of Pakistan. Shortly thereafter, farming expanded further east into the Indus River Valley because of population pressure (Misra 2001: 502). The Neolithic transition in the Indus Valley is linked to the Harappan culture in the literature. Over the course of several thousand years, elements of the Southwest Asian agricultural package eventually migrated eastwards from the Indus Valley into the Ganges Valley and southwards into southern India and Sri Lanka. The expansion of agriculture out of the Indus Valley appears to be linked with arid conditions that suddenly arose in this region roughly 4,000 years ago (see Pokharia et al. 2017).

Figure 8.3. The Indus and Ganges Rivers.



Haplogroup G-M201 records a faint but important signal of the Neolithic transition in South Asia. Sengupta et al. (2006) report that G-M201 represents less than 1 percent of the population in India and for Pakistan frequency is about 5 percent. Among Indo-European-speaking populations, G-M201 attains it highest frequency among Iranian-speaking Kalash and Pashtuns in Pakistan (see **Supplementary Table 8.1**).

Section 6. The Dravidian Languages of South Asia.

The Dravidian language family consists of 85 languages spoken by around 228 million people (*Ethnologue* 2017). Examples include Tamil, Malayalam, Kannada, and Telugu. Dravidian languages are mostly found in southern India. Interestingly, G-M201 mutations attain a frequency of about 5 percent in the Tamil Nadu region at the southern tip of India. This was reported in a study presenting data for over 1,600 men, many of whom are Dravidian-speaking farmers (Arunkumar et al. 2012). Moreover, among the Brahui people, the G-M201 haplogroup attains a frequency of about 16 percent (Di Cristofaro et al. 2013). This population represents a distant linguistic island of two million Dravidian speakers in Pakistan.

The G-M201 data detailed above, along with archaeological and climatological perspectives, suggest that Dravidian expanded from Pakistan to southern India roughly 4,000 years ago. This model of the Dravidian expansion posits the Indus Valley as the putative homeland of Dravidian languages. The Brahui are potential descendants of hunter-gatherers that admixed with Southwest Asian farmers during the Neolithic. They adopted agriculture as a subsistence strategy while retaining Dravidian. With the onset of arid conditions, Dravidian farmers migrated out of the Indus Valley along with Indo-European speaking populations. The Indo-Europeans migrated east into the Ganges Valles. The Dravidians migrated south towards Tamil Nadu.

The discussion of linguistic diversity in South Asia continues in the next chapter (Chapter 9) with the presentation of haplogroup H-M2713 data. Additionally, variants of the J2-M172 mutation provide a stronger signal that traces the origins of the South Asian Neolithic to the Levant (see Chapter 11).

Section 7. The Iranian Languages of Central Asia.

Central Asia consists of Afghanistan, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan. A synthesis of linguistic, archaeological, and genetic perspectives suggests that Indo-Iranian languages and agriculture co-expanded from Iran into Central Asia during the Neolithic. Linguistic relics of this expansion include Farsi, Balochi, Kurdish, Tajiki, and Pashto. Archeological support for this position comes from Jeitun in Turkmenistan. This site represents the earliest expansion of agriculture from Southwest Asia into Central Asia, which occurred about 8,000 years ago (Bellwood 2005: 84-86). Soon thereafter, agriculture appeared among the Hissar culture of Tadjikistan, at Kel'teminar in Kazakhstan, within the Ferghana Valley of Uzbekistan, and at Oshkona in Tadjikistan (Fuller 2006). The expansion of agriculture into Central Asia was facilitated by the so-called Inner Asian Mountain Corridor (e.g., Yatoo et al. 2020; Stevens et al. 2016, Spengler et al 2014). This corridor connects Southwest Asia with Central Asia with a route through the Himalayan mountains and around the Tarim Basin.

Turning now to the genetic data, the frequency of G-M201 among Iranian speaking Pashtuns and Tajiks in Afghanistan suggests that this marker records the history of Indo-European expansions into Central Asia during the Neolithic (see **Supplementary Table 8.1** and data for Central Asia). It should be noted that among the Pashtuns the G2b-M377 mutation stands as a particularly strong G-M201 variant (see **Supplementary Table 8.10**). Furthermore, in Iran, which represents a transit point for the expansion of Indo-Iranian populations into Central and South Asia, the G-M201 haplogroup attains a frequency of almost 12 percent (Grugni et al 2012). These data support the position that Indo-

Iranian populations in Central Asia may have descended from Neolithic farmers who constructed Neolithic farming settlements at Jeitun and further south in the Indus River Valley. Taking this a step further, the Iranian and Indo-Aryan sub-branches represent subsequent regional diversification of the Indo-Iranian language spoken by these farmers.

Figure 8.4. Jeitun.

Section 8. The Turkic Languages of Central Asia.

The Eurasian steppe, a vast biome characterized by grassland and aridity, extends across much of Central Asia. Because of the arid conditions, cereal agriculture is not the ideal subsistence strategy for this region. Rather, the more sustainable food economy is one based on pastoralism. The source of domesticated cattle, goats, and sheep in this region appears to be the Fertile Crescent. This conclusion is based on analysis of animal remains uncovered from the Jeitun archeological site which are dated to about 8,000 ago (Harris 1997). As previously detailed in Chapter 7, the agricultural transition in Central Asia continues with the domestication of the horse roughly 5,500 thousand years ago in north-central Kazakhstan.

In Central Asia, the highest G-M201 frequencies are found among Kazakhs, a Turkic speaking population (see **Supplementary Table 8.1**). One interesting observation stemming from the Kazakhs data is the elevated frequency of the relatively rare G1-M285 mutation among this population (see **Supplementary Table 8.2**). Its presence among the Kazakhs of Central Asia may well represent language shift among populations that had once spoken Indo-Iranian languages. Such a conclusion was reached by Balanovsky et al. (2015), a study that focused on the distribution of G1-M285 variation in Asia. Moreover, based on an analysis of the data, they suggest that the expansion of Indo-Iranian languages correlates well with an expansion of agriculture from Southwest Asia.

As shown by **Supplementary Table 8.2**, the rare G1-M285 mutation also attains a significant frequency among Armenians. Martirosyan (2013) examined lexical correspondences as found among Greek, Armenian, and Indo-Iranian language. He offers a “preliminarily” conclusion and suggests that at the time of the Indo-European dispersal, Armenian, Greek, and Indo-Iranian may have been part of a dialect group. This conclusion helps to interpret that G1-M285 data with the idea that Iranian languages once held a much larger corner of the linguistic tapestry of Central Asia.

Section 9. Haplogroup G-M201, the Caucasus, and Language Contact.

9.1. Overview.

This section will focus on the North Caucasian and Kartvelian language families, as well as Ossetic, an East Iranian language, because they best illustrate the potential of haplogroup G-M201 data for exploring language variation in the Caucasus. The Caucasus region is located between the Black and Caspian Seas, and includes parts of Russia as well as Armenia, Azerbaijan, and Georgia. Within this compact region four different language families are represented: Indo-European, North Caucasian, Kartvelian, and Turkic. Deciphering the complex pattern of language variation in this region involves analysis of genetic, linguistic, historical, and archaeological data to decipher language maintenance and language shift. As such, the region provides opportunities to explore language contact theory. For more details about this approach to language variation, Thomason (2001) and Winford (2003) are recommended.

9.2. Haplogroup G-M201 and the Neolithic Transformation.

According Rootsi et al. (2012), the most common G-M201 variants observed among the contemporary populations of the Caucasus are the G2a-P16 and the G2a-U1 mutations (see, also **Supplementary Tables 8.11 and 8.12**). According to the archeological record, the Southwest Asian agricultural package arrived in the Caucasus region about 8,000 years ago (see Chapter 11 for more details). Interpretation of contemporary Y-chromosome data suggests that the Neolithic transformation in the region also brought farmers from the Fertile Crescent with the haplogroup G-M201 mutation (e.g., Herrera et al. 2012; Rootsi et al 2012; Yunusbayev et al. 2012; Hovhannisyan et al. 2014; Karafet et al. 2016). This model is consistent with ancient Y-chromosome data see **Supplementary Table 8.4**. The oldest G-M201 remains, which date to about 10 thousand years ago, were found at the Boncuklu in south-central Turkey (Kilinc et al. 2016). This archaeological site is located near Çatalhöyü, one of the earliest farming settlements of the Middle East. Thus, the high frequency of G2a-P16 and the G2a-U1 in the Caucasus is linked to the adoption of agriculture in the region and with that, the associated phenomenon of rapid population growth.

9.3. Haplogroup G-M201 and Endogamy.

A review of the G-M201 data reveals an interesting distribution pattern in the Caucasus for this haplogroup. Among several different populations, the haplogroup attains astonishingly high frequency (see **Supplemental Table 8.1**). Moreover, this observation occurs cross-linguistically. For example, among Iranian-speaking Ossetians the frequency is about 70 percent. Among Georgians (Kartvelian languages) the frequency is around 50 percent. A similar frequency is reported for the Abkhaz (North Caucasian languages). Finally, among Turkic-speaking Balkars and Karachays the frequency is also significant, close to 30 percent.

In an interesting paper from 2008, the linguist Bernard Comrie suggests that topography and strict adherence to endogamy (marriage within the same group) explains, at least partially, the extreme linguistic variation found in the Caucasus. From a genetics perspective, topography and endogamy suggest that genetic drift has leveled Y-chromosome variation in the region, producing the high frequency of G-M201 mutations reported in several populations. Taking this a step further, cultural and geographic isolation may explain the prehistory of several language families in the world including North Caucasian and Kartvelian.

9.4. North Caucasian and Kartvelian.

The near or total absence of North Caucasian and Kartvelian languages outside the Caucasus suggests that both language families are indigenous to the region. Moreover, the haplogroup G-M201

Figure 8.5. The Caucasus.

data suggest that both language families represent language maintenance. In other words, the historical incursion of Turkic into the Caucasus (see Johanson 2006) only achieved partial language shift within the region. The idea that North Caucasian and Kartvelian form an indigenous component within the complex pattern of language variation in the Caucasus is supported by the discovery of petroglyphs in the Gorbustan region of Azerbaijan. According to UNESCO, these petroglyphs date to about 12 thousand years ago. As such, human inhabited the Caucasus well before the arrival of agriculture. Additionally, it should be noted haplogroup J-M304 data also support the position that North Caucasian and Kartvelian languages are an indigenous linguistic component of the Caucasus region. Accordingly, this discussion continues in Chapter 11.

9.5. Ossetians.

Ossetians are an ethnic group found in the Republic of North Ossetia in Russia and directly south in the Republic of South Ossetia in Georgia. A useful English-written source of socio-linguistic information about this population comes from conference notes prepared in 2003 by Leila Dodykhudo, a senior researcher at the Russian Academy of Sciences. According to the researcher, the Alans were among the medieval cultures of the Eurasian steppes. In 372, they were defeated by the Huns and some of the Alans eventually fled into the Caucasus. As such, the historical record explains a linguistic peculiarity of the Ossetic language, which is classified as Indo-European, Indo-Iranian, Iranian, and East Iranian. The other East Iranian languages are clustered two thousand kilometers away in Central Asia, mostly in Tajikistan and Afghanistan.

Turning now to the Y-chromosome data, the high frequency of the G2a-P16 mutation observed among the Ossetians, and the absence of the mutation outside the Caucasus (see **Supplementary Table 8.11**), links the Ossetians with the Neolithic transition in the Caucasus. Additionally, haplogroups R-M207 and C2-M217 are essentially absent in this population (see Yunusbayev et al. 2012). As such, the genetic history of Ossetians lacks a Central Asian contribution from steppe nomads. This observation, along with the historical record and the linguistic data, point to language contact and prestige motivated language shift as an explanation for the presence of the Ossetic language in the Caucasus. Thus, the genetic and socio-linguistic data for Ossetians underscores a phenomenon observed elsewhere in the world. Language shift does not require population replacement. The Turkish and Hungarian languages provide examples (see Chapters 7 and 15).

Section 10. Conclusions for Haplogroup G-M201.

The contemporary distribution of G-M201 variation is sometimes puzzling. For example, the G2a-L30 variant is found in Judeo Tats, Bagvalal, and Nogais of the Caucasus region (Karafet et al. 2016). Additionally, the same mutation is also found four thousand kilometers away in Flanders (Larmuseau et al. 2014). Given the distances involved, G-M201 must have expanded very rapidly during the Neolithic, and then the expansion terminated suddenly. The distribution of ancient G-M201 mutations is also interesting. They are regularly detected Neolithic remains, both in Europe and Southwest Asia. However, among contemporary populations of Eurasia, the frequency of G-M201 mutations is generally low. An exception, of course, is the Caucasus region where the frequency of G-M201 is massive.

For linguists, haplogroup G-M201 helps to decipher the prehistory of Indo-European, Kartvelian, North Caucasian, and Dravidian languages. From a triangulated Y-chromosome-based perspective, the contemporary distribution of these language families was clearly shaped by agricultural expansions, language shift, and language maintenance.

Chapter 9: Haplogroup H-M2713.

Section 1. Contemporary Distribution of Haplogroup H-M2713.

Haplogroup H-M2713 is a significant marker for deciphering the population history of South Asia, and in particular, Pakistan and India. Sengupta et al. (2006) suggest that this marker attains a frequency of about 26 percent among Indians. Among Pakistanis the frequency is about 6 percent. Additionally, the H-M2713 mutation attains a significant frequency among several Romani populations in Europe. Finally, H-M2713 and its variants are also observed in some populations of the Middle East, Central Asia, and East Asia, where the overall frequency is low. For linguists, haplogroup H-M2713 presents an informative marker for deciphering the prehistory of Indo-Aryan, Iranian, Dravidian, Austro-Asiatic, and Turkic languages. For further information, the reader is directed to **Supplementary Table 9.1**.

Section 2. The Evolutionary History of H-M2713.

The reader is directed to **Supplementary Figure 1.1** from the first chapter. As shown by the figure, HR-M578 and G-M201 are sister clades. H-M2713 evolved from HR-M578 around 50 thousand years ago (Poznik et al. 2016). The reader is now directed to **Supplementary Figure 9.1** which outlines the internal phylogeny of H-M2713 and its informative downstream variants. The internal structure contains two main divisions, H1-M3061 and H2-P96.

To identify where the H-M2713 main haplogroup evolved, some background information needs to be presented. First, according to International Society for Genetic Genealogy (2017), H2-P96 represents a rare mutation found in contemporary Europe, mostly on Sardinia. A study of 1,194 Sardinians (Francalacci et al. 2015) found 7 men with the mutation, a frequency of less than 1 percent. However, ancient DNA from Neolithic sites in Europe and Turkey suggest that H2-P96 achieved a much wider distribution in prehistoric Europe (see **Supplementary Table 9.2**). Secondly, almost all the published data for haplogroup H-M2713 consists of data for the H1a-M69 mutation among South Asian population where the mutation is ubiquitous. Finally, it should be noted that M69 defined the haplogroup H mutation until 2014. Since then, the M69 has been driven deeper into the haplogroup H phylogeny, first with H1-M69, and later H1a-M69.

Poznik et al. (2006) suggest that H1a-M69 evolved around 47 thousand years ago. Given the age of this mutation and its moderate frequency among South Asian populations, one finds widespread consensus among the geneticists that identifies H1a-M69 as an “indigenous” South Asian mutation (e.g., Sahoo et al. 2006; Sengupta et al. 2006; Trivedi et al. 2008; Debnath et al. 2011; Khurana et al. 2014). Because the phylogeny of H1a-M69 has been revised, with M69 downgraded from a main haplogroup to H1a, the question remains if H-M2713 and H1-M3061 are also indigenous South Asian mutations. Based on the dating estimate and the distribution of Haplogroup H-M2713 mutation, both ancient and contemporary, haplogroup H-M2713 could have evolved in the Levant. H1a-P96 may represent diversification of H-M2713 mutations in South Asia. As such, H1a-P96 is a genetic relic of a human expansion to India via the southern route during Marine Isotope Stage 3, which is supported by the archeological record and paleoclimatological data (see Chapter 4). H2-P96, on the other hand, stands as a relic of haplogroup H-2713 diversification in the Levant among population that remained in the region until the Neolithic, a model that is supported by ancient Y-chromosome data (see **Supplementary Table 9.2**). A recent study (Rohrlach et al. 2021) supports this interpretation of the H2-P96 data by using a new technique called Y-mappable capture assay, which helps to amplify and reconstruct damaged sections of ancient DNA samples.

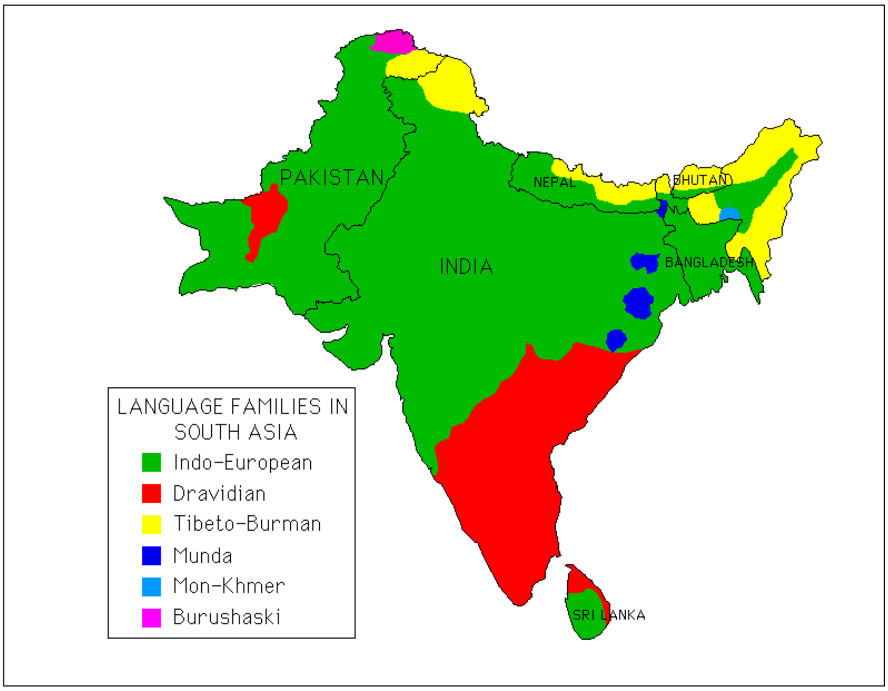
As noted previously, H1a-M69 appears to have evolved about 47 thousand years ago. As such, this mutation represents a Paleolithic component of the South Asian gene pool. Turning now to downstream variants of H1a-M69, three mutations are commonly reported in published reports: H1a1-M52, H1a1a-M82, and H1a2a-Apt. Sengupta et al. (2006) and Karmin et al. (2015) suggest that H1a1-M52 and H1a1a-M82 evolved during the South Asian Neolithic. Dating estimates for H1a2a-Apt, on the other hand, suggest that this mutation evolved during the South Asian Mesolithic (Sengupta et al. 2006).

Section 3. H1a-M69 and Language Variation in South Asia.

3.1. Overview.

For the purposes of this present discussion, the term "South Asia" presents an overview of linguistic variation in Pakistan and India. The linguistic diversity of this region is, indeed, remarkable. In India almost all of the spoken languages fall within one of the four language groups: Dravidian, Indo-European, Austro-Asiatic or Sino-Tibetan. In neighboring Pakistan, on the other hand, Austro-Asiatic is absent. With respect to the Indo-European language family, one main difference between Indian and Pakistani linguistic diversity is that the Indo-European languages of India fall almost exclusively within the Indo-Aryan branch. Hindi, one of India's official languages, and a linguistic heavyweight with over 500 million speakers, provides an example (*Ethnologue* 2018). The Indo-European languages of Pakistan, on the other hand, are a mixture of Iranian and Indo-Aryan. Significant Indo-Aryan languages of this country include Urdu, the official language, and Punjabi, the most widely spoken language (CIA World Fact Book 2018). Within the Iranian branch, Pashto and Balochi are widely spoken.

Figure 9.1. Language Families in South Asia. Source: University of Buffalo, Department of Linguistics.



Turning now to Dravidian languages, 79 languages fall within this classification (Glottolog Version 4.4). Brahui is an "outlier" Dravidian language found in Pakistan. The remaining Dravidian languages, such as Tamil, Telugu, and Kannada, are spoken in southern India. Interestingly, the spatial distribution of Indo-European and Dravidian languages of India generally follows a geographic pattern with Indo-European in the north and Dravidian in the south.

The Sino-Tibetan languages of South Asia fall within the Tibeto-Burman branch. Within India, the distribution of Tibeto-Burman languages is found along the border that this country shares with Nepal and China. It should be noted that an attempt was made to extrapolate the number of Tibeto-Burman languages spoken in India from the *Ethnologue* website. This proved difficult, but the figure appears to be around 125 languages. Examples include Mizo, a language spoken by around 675 thousand people (Ethnologue 2018). Interestingly, only a single Tibeto-Burman language is listed for Pakistan, Balti with around 327 thousand speakers.

According to Ethnologue (2018), the Austro-Asiatic language family consists of 167 languages. These languages stretch along a geographical expanse than begins in eastern India and end in Malaysia. Within this language family, the Munda branch represents almost all of the Austro-Asiatic languages of eastern India. Santhali and Mundari are among the more recognized Munda languages. The Mon-Khmer branch, on the other hand, represents the Austro-Asiatic languages spoken in East Asia. Significant Mon-Khmer languages of the region include Khmer and Vietnamese.

3.2. South Asian H1a-M69 Data and Ascertainment Bias.

Two studies, Sengupta et al. (2006) and Trivedi et al. (2008), presented frequency data that facilitate analysis of the extent to which haplogroup H1a-M69 is an informative mutation among the four main language families of India. The problem with the studies is that the sample sizes are small and as such, ascertainment bias becomes a problem. In order to overcome this problem, **Supplementary Tables 9.3 through 9.6** explore the correlation between linguistic and genetic diversity in South Asia by utilizing a larger dataset of over 7,000 samples. The tables were prepared in order to determine the frequency of H1a-M69 in South Asians more accurately according to language family or language branch. In order to minimize ascertainment bias, the tables excluded data from populations for which the sample size was less than 20 men. The data were then compared against the results obtained by Sengupta et al. (2006) and Trivedi et al. (2008), which are summarized in Table 8.1 (below).

Table 9.1. South Asian Languages and H1a-M69.

	Sengupta et al. 2006 n = 709	Trivedi et al. 2007 n = 1,152	Present Report n = 7,551
Frequency of H1a-M69 among Indo-European populations.	28%	26%	17%
Frequency of H1a-M69 among Dravidian populations.	33%	30%	28%
Frequency of H1a-M69 among Austro-Asiatic populations.	23%	6%	25%
Frequency of H1a-M69 among Tibeto-Burman populations.	2%	10%	6%

The data presented in the above table confirm that the H1a-M69 mutation is a significant marker among Indo-European, Dravidian, and Austro-Asiatic speaking populations in South Asia. Among Tibeto-Burman speaking population, the mutation attains a low frequency. It should be noted that the frequency for Tibeto-Burman is probably over-inflated because the H1a-M69 mutation was not detected in several of the reported Tibeto-Burman-speaking populations, and these populations are not included in the analysis. Indo-European, Dravidian, and Austro-Asiatic populations, on the other hand, almost always have the H1a-M69 mutation. Consequently, the overall frequency data for these three language families are more accurate than that reported for Tibeto-Burman.

3.3. H1a-M69, the Neolithic, Dravidian, and Indo-European.

As previously discussion in Chapter 6, deciphering the linguistic prehistory of South Asia requires researchers to decipher Paleolithic, Mesolithic, and Neolithic components of the genome within this region. The same chapter defines the C1b-M356 mutation as a Paleolithic component. In Chapter 8, haplogroup G-M201 was identified as a Neolithic component. This present chapter (Chapter 9) defines the H1a-M69 as a Paleolithic component. This knowledge will facilitate the interpretation of data from other mutations found in the region, namely J2-M172, L-M20, T-M184, O-M175, and R1a-Z93 (see Chapters 11, 12, 16, and 18).

The reader is directed once again to **Supplementary Table 9.1**. What is particularly striking about the Pakistani data is that the H1a-M69 mutation appears among the Indo-Aryan-speaking Kalash and Iranian-speaking Pathans. The mutation also appears among the Dravidian-speaking Brahui people and among the Burusho, speakers of a language isolate. Unfortunately, the amount of Pakistani data is small, but they suggest nevertheless that H1a-M69 may well have been part of genetic inventory of hunter-gatherers who lived in the Balochistan region when Indo-European-speaking farmers arrived about nine thousand years ago. This suggests that the Neolithic in South Asia involved the admixture of Southwest Asian farmers and South Asian hunter-gatherers.

As previously mentioned in Chapter 8, the expansion of Indo-European and Dravidian across South Asia is linked with farmers from Southwest Asia who settled at Mehrgarh in the Balochistan region of Pakistan around 9,000 years ago. Over the course of several thousand years, this agricultural trajectory penetrated the Indus River Valley of Pakistan and western India. From this location, agriculture eventually expanded further eastwards into the Ganges River Valley, and southwards into southern India and Sri Lanka. From an overall general perspective, farmers who expanded into the Ganges River Valley now speak Indo-Aryan, and those who expanded southwards now speak Dravidian. Taking this a step further, agricultural expansions from a common region, language maintenance, and language shift, all converge to explain why H1a-M69 mutations attain a significant frequency among populations that speak Dravidian and Indo-Aryan.

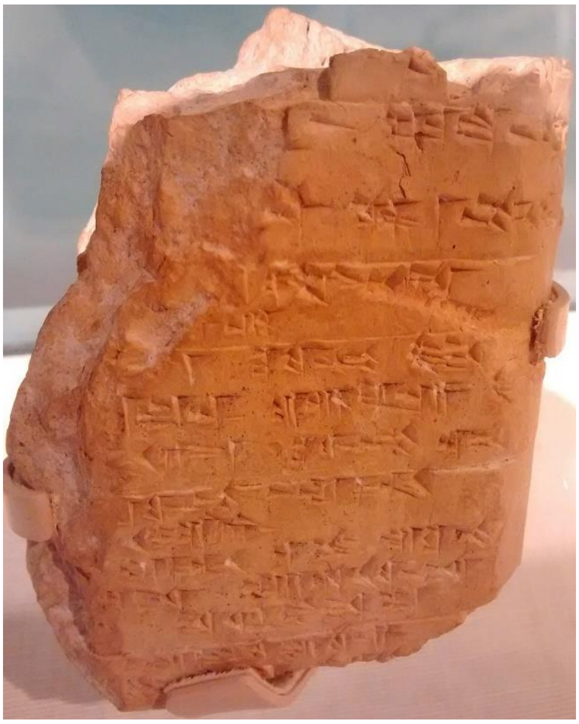
3.4. H1a-M69, the East Asian Neolithic, and Austro-Asiatic.

Based on the available data, Dravidian stands as a relic of Mesolithic hunters-gatherers who adopted agriculture but resisted language shift to Indo-Aryan or Iranian. The Indo-European languages of South Asia, on the other hand, are a linguistic relic of Mesolithic hunter-gatherers who adopted the Southwest Asian agriculture package and shifted languages. Finally, Austro-Asiatic stands as a relic of the East Asian Neolithic and the expansion of rice agriculture from China into eastern India about 4,000 years ago, a topic that will be discussed in greater detail in Chapter 16. At this point, it is important to suggest the following: Haplogroup H1a-M69 mutations observed among Austro-Asiatic populations in India represent a collision of Southwest Asian and East Asian agricultural expansions and language shift from Dravidian or Indo-European to Austro-Asiatic.

Section 4. H1a-M69 and Language Variation in Central Asia.

H1a-M69 variation in South Asia potentially offers useful data for assessing the purported Central Asian origins of Indo-Iranian languages and the evolution of the Iranian and Indo-Aryan branches. Such a discussion requires the presentation of important background information about the origins of the so-called Central Asian *steppe nomad hypothesis*. It should be noted that at the beginning of the twentieth century, with the discovery of clay tablets at Boğazkale in modern-day Turkey, Hittite became the oldest attested Indo-European language. During the nineteenth century, however, Sanskrit was considered the oldest attested Indo-European language. For this reason, many linguists, such as Max Müller, took a keen interest in Sanskrit and the Aryan lords mentioned in the Rigveda liturgical texts (for a more detailed discussion see Pedersen 1967 and Arvidsson 2006). From their interpretation (or perhaps misinterpretation) of these texts evolved the idea that the Aryan people were the original speakers of an Indo-European language. During the twentieth century Nazi Germany re-worked the Aryan hypothesis to support their racial and ethnic ideology (for a more detailed discussion see Pringle 2006). The archaeologist Marija Gimbutas then reworked the Aryan hypothesis in a series of articles published between 1952 and 1993. Instead of Aryans she proposed that the first Indo-Europeans were the prehistoric Kurgan people of the Russian steppes. Today, one of the most recognized contemporary proponents of this approach to Indo-European origins is the anthropologist David Anthony (for more details see his 2007 monograph).

Figure 9.2. Hittite Cuneiform Inscription. Source: Wikipedia and Mx. Granger.



Contemporary discussions of Indo-European origins have re-worked Gimbutas' Kurgans hypothesis into a Central Asian steppe nomad hypothesis. Several Y-chromosome studies have proposed or rejected the potential contribution of steppe nomads to the South Asian gene pool (e.g., Kivisild et al. 2003; Cordaux et al. 2004; Sahoo et al. 2006; Sengupta et al. 2006; Trivedi et al. 2008). While final analysis of the Central Asian steppe nomad hypothesis must wait until Chapter 18 and the discussion of the R1a-Z93 mutation, it is necessary at this time to present the H1a-M69 perspective.

Interestingly, the H1a-M69 mutation has been detected in several populations of the Central Asia and even in East Asia (see **Supplementary Table 9.1** for additional information). These populations include those that speak Iranian languages, such as Tajiks and Pashtuns. The mutation has also been detected in Turkic-speaking populations, such as Uygur and Uzbeks. South to north geneflow could explain the source of H1a-M69 mutations in Central and East Asia. Archaeological support for this position stems from the Inner Asian Mountain Corridor, a topic that appears in Chapter 11 and the discussion of haplogroup J-M304. Perhaps the prehistoric Tocharian people of the Tarim Basin in the Xingjian region of China also stand as a linguistic relic of this trade network. They spoke an Indo-European language branch that became extinct about a thousand years ago.

Section 5. H1a-M69 and the Romani Languages.

The H1a-M69 mutation has also surfaced as a useful marker for deciphering the population history of the Romani people, who are often identified as Roma, and sometimes as Gypsies, a term that is considered derogatory. This population is found throughout Europe. For years scholars have asked whether India is the putative homeland of these people, a conclusion supported by the linguistic data. The Romani language is classified as an Indo-Aryan language. Further support also comes from the historical record (e.g., Tcherenkov and Laederich 2004). Finally, the Y-chromosome data identify India as the putative homeland of the Romani. The H1a-M69 mutation attains a significant frequency among many of the Roma groups in Europe, such as 17% among the Iberian Roma (Gusmão et al. 2008), and 32% among the Hungarian Roma (Pamjav et al. 2011). Moreover, a 2012 study published by Rai et al. 2012 analyzed haplogroup H1a-M69 data that was taken from 10 thousand global samples. Based on their analysis, they identified northeast India as the putative homeland of the Romani people, a conclusion that is consistent with the linguistic and historical data.

Section 6. The FR-M89 Paragroup.

It should be noted that haplogroup H-M2713 frequencies for South Asia might be underreported in the literature. ISOGG 2017 states that when M69 was the main haplogroup H mutation, potential H-M2713 and H1-M3061 mutations for South Asia were identified as unspecified variants of the FR-M89 paragroup (e.g., Cordaux et al. 2004; Sengupta et al. 2006; Arunkumar et al. 2012; Khurana et al. 2014). Clearly, further testing is needed to attain a more accurate determination of haplogroup H-M2713 variation in the region.

The actual FR-M89 paragroup, and not H-M2713, was detected among the Soliga people, a Dravidian-speaking tribe residing at the southernmost tip of India (see Rowald et al. 2020). Their presence at this location, along with their Negrito-like appearance and traditional hunter-gatherer subsistence strategy, suggest that they are descendants of the out-of-Africa migration. The significant frequency of the C1b-M356 mutation among these people also supports such a conclusion. However, the H1a-M82, J2b-M241, R2a-M124, and L1a-M27 mutations, along with the linguistic data, suggest admixture with Dravidian people who arrived in southern India during the Neolithic.

Rowold et al. (2020), based on a comparison of the short tandem repeat (STR) data for Soliga and that from Africa, suggest a South Asia to Africa migration sometime in the prehistoric past. This conclusion is problematic because it is based on insufficient data using an unreliable marker. Rather, the genetic tool of choice is Y-chromosome single nucleotide mutations. As such, a prehistoric migration from India to Europe seems plausible based on ancient H2-P96 remains found in Europe. Nevertheless, as previously noted in Section 2 (above), the currently available data suggest that H2-P96 represents diversification of haplogroup H-2713 among populations that remained in the Levant until the Neolithic.

Section 7. Conclusions for Haplogroup H-M2713.

The discussion of language variation in South Asia continues in Chapter 11 and the presentation of J2-M172 data for the region. Turning now to the present discussion of haplogroup H-M2713, almost all the published frequency data consist of the H1a-M69 mutation and its downstream variants. The currently available data suggest that H1a-M69 is a Paleolithic component of the South Asian genome. For linguists, this marker helps to decipher language variation in this region. Dravidian appears to be an indigenous linguistic component from eastern Pakistan. Indo-European was imported from the Levant. Austro-Asiatic and Tibeto-Burman came from East Asia. As such, the presence of H1a-M69 among Indo-Aryan and Austro-Asiatic populations underscores the influence of language shift and language maintenance as a force that has shaped the complex linguistic tapestry of South Asia.

Chapter 10: Haplogroup I-M170.

Section 1. The Contemporary Distribution of Haplogroup I-M170.

At this point the reader is directed to **Supplementary Table 10.1** which provides a survey of haplogroup I-M170 frequencies across Eurasia. While the haplogroup appears sporadically among populations in the Middle East, the Caucasus, and Central Asia, I-M170 represents a significant marker among European populations where, according to Underhill et al. (2007), roughly 20 percent of men have the mutation. The same study also suggests that I-M170 is the only Y-chromosome haplogroup that evolved on the European continent. The remainder of European Y-chromosome variation (e.g., R1b-343, R1a-M420, J2-M172, E1b-V13, G2a-P15, and N1a-Tat) arose from haplogroups that evolved in Asia.

Published studies provide contemporary frequency data for the following I-M170 variants: I1-M253, I2a-M423, I2a-M26, and I2a-M223. The I1-M253 mutation (see **Supplementary Table 10.2**) attains a significant frequency among the Germanic and Uralic-speaking populations of Scandinavia. Among the South Slavic-speaking populations of the Balkans region of Europe, the I2a-M423 mutation (see **Supplementary Table 10.3**) attains an especially high frequency. Similarly, I2a-M26 (**Supplementary Table 10.4**) attains a significant frequency on the island of Sardinia in the Mediterranean. Finally, I2a-M223 (**Supplementary Table 10.5**), attains low frequency numbers throughout Europe.

Three I-M170 variants are linguistically informative, the I1-M253, I2a-M423, and I2a-M26 mutations. I1-M253 elucidates the prehistory of Germanic and Uralic languages. The same mutation also explains the historical expansion of the Romani languages, which are part of the Indo-Aryan language branch. The I2a-M423 mutation, on the other hand, helps to decipher the historical expansion of Slavic languages. Finally, I2a-M26 supports an argument that defines the Basque language isolate as a linguistic relic of pre-Neolithic Europe.

Section 2. The Evolutionary History of I-M170.

2.1. Overview.

The I-M170 haplogroup and its downstream variants present useful mutation for deciphering the arrival of *Homo sapiens* in Europe, the so-called Aurignacian cultural tradition. Additionally, they are genetic relics of populations expansions from southern European refugia following the Last Glacial Maximum.

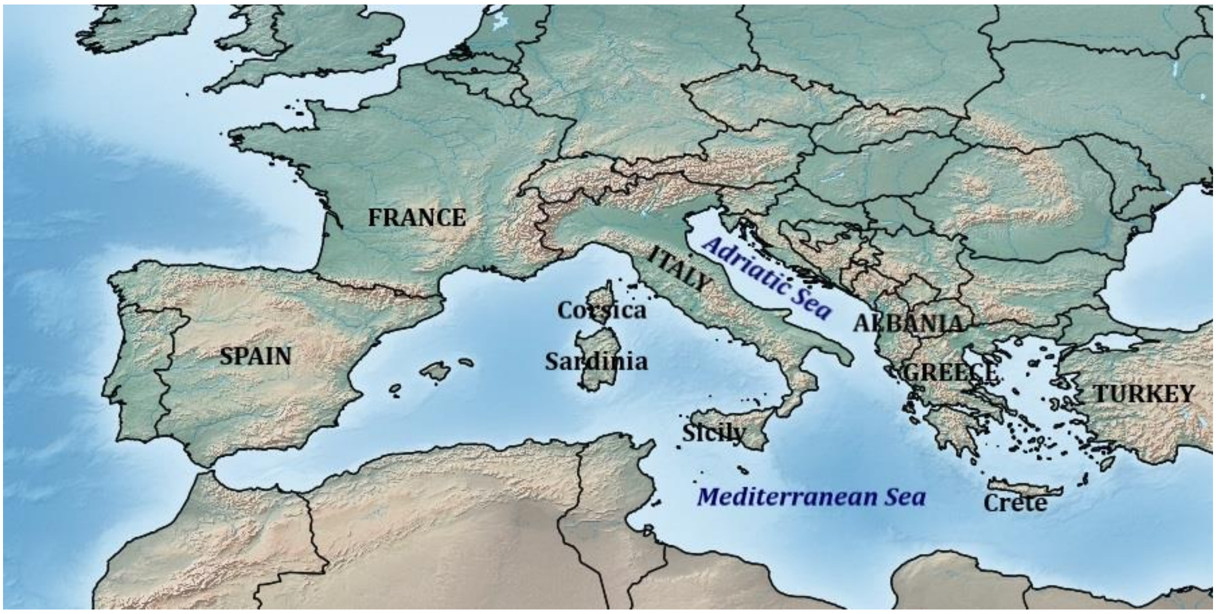
2.2. Important Phylogenetic Relationships.

The reader is directed to **Supplementary Figure 1.1** which depicts the important evolutionary steps between Y-Chromosome Adam and the main haplogroups. According to Poznik et al. (2016), haplogroup I-M170 evolved from the IJ-M429 paragroup about 44 thousand years ago. The reader is now directed to **Supplementary Figure 10.1** which presents the internal phylogeny of haplogroup I-M170. Within this main haplogroup, I1-M253 and I2-M438 represent the two main internal clades. Dating estimates from Underhill et al. (2007) suggest that both mutations separated from I-M170 about 28 thousand years ago, a point in the prehistory that correlates well with the Last Glacial Maximum and the idea that diversification of haplogroup I-M170 occurred in Ice Age refugia.

2.3. Marine Isotope Stage 3.

A discussion of the initial human colonization of Europe during the Paleolithic helps to explain the contemporary distribution of haplogroup I-M170 mutations. Genetic, archeological, and climate previously presented in Chapters 4 and 6 support an expansion of the human tribe during Marine Isotope Stage 3. Turning briefly to the climatological data, as previously detailed, warmer weather during Marine Isotope Stage 3 facilitated an expansion of *Homo sapiens* out of southwestern Asia. Furthermore, climate change facilitated not only the initial colonization of South and East Asia (see Chapter 4), but Europe as well (see Müller et al. 2011).

Figure 10.1. Europe and the Mediterranean.



Important archaeological support for the initial human colonization of Europe comes from Hoffecker (2009) who reports the discovery of artifacts found in Eastern Europe and in Mediterranean region of the continent. These artifacts include spear points and scrappers made from stone. These tools, artifacts of Aurignacian archeological tradition, were constructed about 48 thousand years ago. Important support from the human fossil record comes from the Bacho Kiro Cave in Bulgaria. Two individuals have the C1-F3393 mutation, and one has the F-M89 mutation. Dating results suggest that they died approximately 44 thousand years ago (for more details, see Hublin et al. 2020; Hajdinjak et al. 2021). Another study (Prüfer et al. 2021) report data obtained from a skull found inside the Koněprusy cave system in the Czech Republic. These remains are from as adult female who died roughly 45 thousand years ago. Additional human fossil evidence also comes from the Peştera cu Oase cave in Romania. These remains are dated to about 40 thousand years ago (Trinkaus et al. 2003; and Fu et al. 2014).

Hoffecker (2009), based on an interpretation of the archaeological record, suggests that human colonization of Europe involved a migration route through Central Europe. As previously detailed in Chapter 6, C1-F3393 data supports this model. As previously suggested, this migration was undertaken by Paleolithic hunter-gatherers who once roamed the mammoth steppes in pursuit of large herbivores, such as reindeer, bison, mammoths, and horses. In his 2009 report, Hoffecker also suggests that humans colonized Europe via a second expansion along the Mediterranean Sea. Perhaps hunter-gatherers undertook this migration to harvest marine resources. Important Y-chromosome support for this expansion comes from the Paglicci Cave in southern Italy where the oldest I-M170 remains were found (Fu et al. 2016). The data come from an individual who died about 33 thousand years (for more details, see Supplementary Table 10.6). Important archaeological support for the Mediterranean expansion comes from Aurignacian artifacts recently discovered in Portugal and southern Spain (see Haws et al. 2020). Similar archaeological data also surfaced in a published report from 2019 (Carrión et al.). Both studies date the presence of *Homo sapiens* in southern Iberia to around forty thousand years ago.

2.4. Marine Isotope Stage 2 and Ice Age Refugia.

Haplogroup I-M170 is not only a genetic relic of the human colonization of Europe, but also represents the genetic relic of hunter-gatherer populations that survived the last Ice Age. During Marine Isotope Stage 2, which began about 29 thousand years ago, the weather became colder in Europe. Shortly thereafter, about 26 thousand years ago, the ice glaciers reached their maximum southern expansion across Eurasia. In western Europe, glaciation extended roughly to the Pyrenees Mountains along the Franco-Spanish border. The literature typically refers to this point in time as the Last Glacial Maximum (LGM). This change in climatic conditions drove human populations in Europe to the southern part of the continent to habitable regions commonly defined as refugia.

According to Binney et al. (2016), after the ice glaciers had started to recede around 21 thousand years ago, the European landscape above the fortieth northern parallel became a treeless region of tundra. Around 14 thousand years ago, as the climate warmed, the European tundra also began to contract northwards, leaving behind areas of forest. Around 11 thousand years ago, the tundra reached Scandinavia. Finally, about four thousand years ago, the tundra reached its current location along the Arctic Circle.

Beginning about 14 thousand years ago, close to the beginning of Marine Isotope Stage 1 and the Holocene, warmer weather and retreating ice glaciers allowed human populations to migrate out of the southern European refugia and re-colonize the depopulated regions of Western Europe and Scandinavia. Haplogroup I-M170 support for this model includes remains from southern Spain dated to 24 thousand years ago; I2a-M223 remains from Belgium dated to 15 thousand years ago; I1-M253 remains from northern Spain dated to about 13 thousand years ago; I2-M438 remains from France dated to 13 thousand years ago; I2a-P37.2 remains from Switzerland dated to 14 thousand years ago; and I-M170 remains in Norway dated to 9 thousand years ago (see **Supplementary Tables 10.6** to **10.14**).

Section 3. The I1-M253 Mutation and Germanic Languages.

3.1. Overview.

Interpretation of the linguistic and archaeological data has long identified Denmark as the putative homeland of Germanic languages (e.g., St. Clair 2012). The Y-chromosome data, especially the I1-M253 mutation, also support this conclusion. The frequency of the I-M253 mutation in Scandinavia furthers suggests that this marker is the genetic tool of choice for understanding the linguistic prehistory of Germanic languages. As shown by **Supplementary Table 10.2**, the I1-M253 mutation attains frequencies between 30 and 40 percent among the contemporary populations of the region. These populations include North Germanic-speaking Danes, Swedes and Norwegians and Uralic-speaking Finns and Sami. Elsewhere in Europe, I-M253 attains a moderate frequency among the Dutch, British, Flemish, and Germans, speakers of West Germanic.

3.2. Prehistory of Germanic Languages from a Linguistic Perspective.

The role of language contact theory in shaping early Germanic languages was explored by St. Clair in his 2012 Ph.D. dissertation at the University of California. One controversial idea stemming from the dissertation is that Germanic languages evolved as the result of language contact between speakers of proto-Basque, proto-Indo-European and proto-Afro-Asiatic. Perhaps less controversial would be the idea that Germanic languages have considerable time depth. The prehistoric evolution of Germanic potentially reflects a long-term process over a ten-thousand-year period that included human adaptation to climate change, the isolation of populations from each other, and the Neolithic transformation.

From a linguistic perspective, the mixed origins of Germanic are plausible. Linguists have long noted that perhaps a third of the Modern German lexicon lacks an Indo-European cognate (e.g., Vennemann 2000: 241; Waterman 1976: 36; Schirmer and Mitzka 1969). Additionally, Mailhammer (2007) suggests that the systematic pattern of ablaut for Germanic strong verbs may have been a featured borrowed from Afro-Asiatic languages. Interestingly, the presence of proto-Afro-Asiatic languages in prehistoric Europe is plausible from a genetic perspective. As previously suggested in Chapter 5, the E1b-V13 mutation may signal the expansion of proto-Afro-Asiatic into Europe during the Mesolithic. However, the mutation is mainly found in the Mediterranean and Balkans and is virtually absent in Scandinavia. Thus, evidence of prehistoric contact between Scandinavia and southeastern Europe requires additional research.

3.3. Prehistory of Germanic Languages and Archaeology.

Figure 10.2. Photo of a Norwegian Reindeer. Source: Wikipedia and Are G Nilsen.

Tundra is the preferred habitat of reindeer as they are especially fond of lichen that thrive in this biome. Among prehistoric *Homo sapiens* in Eurasia, reindeer meat was an important source of food. The prehistoric foragers who hunted this animal represent a key component in understanding contemporary language variation across Eurasia. Focusing now on Scandinavia, Siiräinen (2003) suggests that the human colonization of northern Europe during the Holocene was facilitated northward contraction of the ice glaciers, tundra, and reindeer. As the tundra began to recede, about 14 thousand years ago, large herds of reindeer migrated northwards into central Europe. The reindeer eventually reached Scandinavia about 12 thousand years ago. Close behind were people that hunted these animals, the so-called Ahrensburg culture, who eventually settled in the region.

Around 10 thousand years the landscape in Scandinavia transitioned from tundra to forests. This transition forced a change in subsistence strategy because the reindeer disappeared into the Arctic Circle. As the result of climate change, people in the region became dependent on marine resources, such as mussels (e.g., Lewis et al. 2016). Inland resources, such as elk, were also an important source of food (e.g., Jessen 2015).

As detailed in Chapter 8 and the discussion of haplogroup G-M201, agriculture and the Linear Pottery culture expanded across Europe during the Neolithic. The same expansion disseminated Indo-European languages throughout the continent. When this expansion terminated at the Northern European coastal plain, about 7,500 thousand years ago, agriculture was not embraced by the Mesolithic hunter-gatherers of Scandinavia. Rather, the terminal point of the Linear Pottery expansion became a cultural boundary that lasted about two thousand years until the Funnel-Beaker cultural transition.

Figure 10.3. Tundra. Source: Wikipedia and Dr. Andreas Hugentobler.

The reasons for the slow transition to agriculture in Scandinavia remain a mystery (e.g., St. Clair 2012). One possible explanation assumes that the Linear Pottery culture expansion probably carried agriculture though areas of central Europe which were uninhabited or sparsely inhabited by nomadic foragers. In contrast to central Europe, the Mesolithic peoples of contemporary Denmark lived in permanent or semi-permanent settlements. As such the region probably had a larger population density relative to that of central Europe. Conditions were different in Mesolithic Denmark because of the abundance of marine resources, and with that, the availability of a year-round source of very nutritious food. In other words, the food supply remained stationary, and the land supported more people per square kilometer. Taking this a step further, Mesolithic Scandinavians did not need agriculture.

3.4. Germanic Languages and the I1-M253 Mutation.

The Neolithic began in Scandinavia around five thousand years ago. Three different models have surfaced for explaining this transition: human migration, a food shortage, or socio-economic change (Fischer 2002). The idea of human migration deserves particular attention because it undermines or supports the role of language contact theory in shaping the evolution of Germanic languages. Genetic data can be a useful tool for resolving this question as these data can identify Paleolithic, Mesolithic, and Neolithic components of the contemporary Scandinavian gene pool.

It should be emphasized once again that the I1-M253 mutation attains a substantial frequency among contemporary Scandinavian populations (see **Supplementary Table 10.2**). With this in mind, the reader's attention is now directed to **Supplementary Table 10.14** and the survey of ancient I-M170 mutations found in prehistoric Scandinavia. As shown by the table, the I1-M253 mutations is completely absent from the ancient Y-chromosome data for region. Rather, it would appear that prehistoric Scandinavians had the I2a-M423 mutation. As such, some may argue that I1-M253 was not included in the genetic inventory of prehistoric Scandinavia.

A salient point for understanding ancient and contemporary genetic variation in Scandinavia emphasizes that the present-day land crossing from central Europe into the region has just one route. It involves a journey through Denmark and across the Öresund Straight via bridge and tunnel into Sweden. However, as explained and illustrated by Sporrong (2003), the landscape in Scandinavian was far different 12 thousand years ago. The water level was around much lower, and a larger landmass connected central Europe with Scandinavia. Additionally, the present-day Baltic Sea was a smaller freshwater lake, the so-called Baltic Ice Lake. Consequently, during the early Holocene several different routes presented an opportunity for human settlement in Scandinavia via a land crossing or a short water crossing. The Mesolithic founding populations of Scandinavia probably had I1-M253 and I2a-M423. As suggested by Underhill et al. (2007), I1-M253 entered Denmark via northwestern Europe. Hunter-gatherers with I2a-M423 mutation probably entered Scandinavia through another route further east.

Figure 10.4. Europe and the Baltic Sea.

The demise of I2a-M423 and predominance of I1-M253 in contemporary Scandinavia is probably the result of demographic and cultural developments that occurred in Scandinavia after the Neolithic transition in this region, beginning roughly 5,000 years ago. Agriculture was first adopted in Denmark, and over time the technology spread northwards into the rest of Scandinavia (e.g., Siiräinen 2003). Rapid population growth occurred because agriculture supports higher population density. Thus, a sudden and rapid increase of men with the I1-M253 mutation in Denmark, beginning about 4,000 years ago, and their subsequent migration into northern Scandinavia, probably changed the distribution and frequency of I-M170 variation in the region. Later, the historical expansions of the Germanic tribes brought the I1-M253 mutations to Germany, the Netherlands, the British Isles, and elsewhere in Europe. Dutch, German, Flemish and English represent linguistic relics of this expansion, and with that, the evolution of West Germanic languages.

The above model of haplogroup I-M170 variation in Scandinavia is consistent with a 2006 study (Karlsson et al.) that analyzed almost four hundred DNA samples collected from men in Sweden. With their analysis, they concluded that the arrival of agriculture in Scandinavia occurred as the result of the adoption of a new technology by people already living in the region. As such, the genetic data fail to associate the Scandinavian Neolithic with an influx of Central European farmers. Their conclusion partially follows the heavy frequency of I1-M253 mutations in the region, and the low frequency of Central European Neolithic markers, namely J2-M172 and G2a-P15.

3.5. Indo-European and Language Contact.

Analysis of genetic, archaeological, climatological, and linguistic data suggests that the origins of Germanic languages also provide valuable insight into the evolution of Indo-European languages in Europe. Traditional linguistic opinion, which currently evolves around the "steppe nomad" hypothesis, posits an invasion of Indo-European-speaking people from Central Asia during the Bronze Age, roughly 3,000 years ago. According to this perspective, their language later diversified into Germanic, Slavic, Celtic, and other language branches. However, an evolving picture of Germanic and other Indo-European languages suggests that the Mesolithic languages of pre-agricultural Europe may have influenced linguistic diversity on the continent. Perhaps Germanic was a non-Indo-European language that became Indo-European after the Neolithic transition in Scandinavia. A similar argument can be formulated for Celtic (see Chapter 18).

Section 4. Uralic and the I1-M253 Mutation.

Within Scandinavia the pattern of language variation consists of North Germanic languages and Uralic languages. The Uralic-speaking people of the regional are Finns and Sami. Among these populations, the N1a-M46 mutation (see Chapter 15) attains a significant frequency. However, this mutation is virtually absent among Danes (Sanchez et al. 2004). Among the Norwegians, less than 3 percent have the mutation (Dupuy et al. 2006). Among Swedes, the figure stands between 10 and 14 percent (Karlsson et al. 2006; Lappalainen et al. 2006). The I1a-M253 mutation, on the other hand, is distributed more evenly among all the Scandinavian populations. It is present in about one third of Finns and Saami (Tambets et al. 2004; Lappalainen et al. 2006). Similar frequencies are detected among Danes, Norwegians, and Swedes (see **Supplementary Table 10.2**).

The arrival of Uralic languages in Scandinavia is associated with the Comb Ceramic Culture. The archeological record suggests that this occurred around 5,000 years ago (e.g., Siiräinen 2003). The genetic relic of this culture is the N1a-M46, a mutation that will be discussed in much greater detail in Chapter 15.

Section 5. Romani and the I1-M253 Mutation.

As explained in Chapter 8, the H1a-M69 mutation and the historical record point to India as the putative homeland of the Romani people. Surprisingly, several studies report an unexpected frequency of I1-M253 found among Romani populations in Europe (Gusmão et al. 2008; Petrejčíková et al. 2009; Pamjav et al. 2011. This is surprising because haplogroups I-M170 mutations are virtually absent in South Asia.

The route undertaken by Romani during their migration from India to Europe involved an extended "stopover" in the Byzantine Empire (Tcherenkov and Laederich 2004). St. Clair (2014) suggests that during this stopover, cultural exchange and genetic admixture may have taken place between Romani and the Crimean Goths who also part of the empire. This potentially explains how the Romani acquired metal working skill and I1-M253 mutations. When the Byzantine Empire collapsed, the Romani expanded into Europe with their culture and genes.

Section 6. Linguistic Significance of the I-M423 Mutation.

As previously detailed above, I1-M253 and I2-M438 evolved roughly 28 thousand years ago. The mutations evolved close to a point in the prehistory that is defined by the Last Glacial Maximum. As such, the evolution of both mutations reflects genetic diversification that occurred when Paleolithic populations retreated to southern European refugia at the onset of Marine Isotope Stage 2.

The I2a-M423 mutation is a downstream variant of I2-M438. Interpretation of the contemporary data suggests that the mutation evolved roughly 9,000 years ago in the Balkans region (Rootsi et al. 2004; Underhill et al. 2007). Interestingly, the ancient Y-chromosome data open the possibility that I2a-M423 evolved in Western Europe, perhaps in Spain or France, rather than in Balkans (see **Supplementary Table 10.11**).

The reader is now directed to **Supplementary Table 10.3**, which provides a survey of the I2a-M423 mutation among contemporary populations. This mutation attains an especially high frequency among several populations of the Balkans region of Eastern Europe, such as Serbs, Croats, and Bosnians. More moderate frequencies are reported for other populations in Eastern Europe, including Moldavians, Bulgarians, Ukrainians, Czechs, and Slovenians. For linguists, the frequency of I2a-M423 mutation among Slavic-speaking populations raises an interesting question, whether the historical expansion of Slavic languages is associated with a large-scale population expansion.

Two useful studies (Sarac et al. 2016 and Karachanak et al. 2013) explore contemporary Y-chromosome diversity among Croatians and Bulgarians. Both studies suggest that the genome of contemporary Slavic-speaking Europeans has Paleolithic, Mesolithic, and Neolithic components. The I2a-M423, R1b-343, and R1a-M420 mutations represents a Paleolithic component. E1a-V13 is a potential Mesolithic relic. Neolithic farmers from Southwest Asia contributed J2-M172 and G2a-P15. Based on these data, the Neolithic transition in Eastern Europe involved a farming expansion that originated outside the region and adoption of a new technology by hunter-gatherers already living in the region.

Brackney (2007) presents a discussion of Slavic origins from historical and linguistic perspectives. He suggests that Common Slavic split into East, West, and South variants in the sixth century (2007: 18). He also builds a model of Common Slavic using the language farming theory (Renfrew 1987). Brackney proposes that the first farmers of Europe were also the linguistic ancestors of the Slavic-speaking populations. He then proposes that the farmers settled north of the Carpathian Mountains in central and eastern Europe during the Neolithic. Brackney asserts that the Slavic-speaking farmers remained in relative isolation in this area until the fifth century. During this period the Indo-European language of Neolithic farmers diversified into Common Slavic. The socio-political situation climate in Europe then changed, creating the opportunity for the Slavic tribes to assert their culture and language onto the historical stage (2007: 91-99).

The above discussion of Brackney and his view of Slavic origins, as well the genetic evidence previously presented, especially the dating estimates, fail to link the contemporary distribution of Slavic with a large-scale population expansion. As such, language contact provides the best theoretical approach for explaining the how and when of the Slavic expansion. Perhaps the language and culture of Slavic peoples filled a void caused by the collapse of the Roman Empire.

The R1a-M448 and R1a-M558 mutations are also important markers for deciphering the expansion of Slavic languages. Accordingly, this topic surfaces in Chapter 18.

Section 7. I2a-M26 and the Basque Language Isolate.

Researchers have long suspected that the populations of the Pyrenees are a relic of pre-Neolithic Europe. Partial support for this position stems from the Basque people who speak a language isolate, which is unusual as western European populations speak an Indo-European language. Since Indo-European languages potentially came to Europe during the Neolithic, roughly 8,000 years ago, one could argue that the Basque language isolate is a relic of European linguistic diversity that predates the arrival of Indo-European languages (e.g., Trask 1996: 191; Vennemann 1994: 263)

Among contemporary populations, the I2a-M26 mutation attains a heavy frequency on the Mediterranean island of Sardinia. Elsewhere in Europe the mutation is absent or exhibits a low frequency (see **Supplementary Table 10.4**). An exception to this distribution pattern might be the Basque region which straddles the contemporary border between France and Spain. Rootsi et al. (2004) reports that about 6 percent of the Basques have the I2a-M26 mutation. Another report, López-Parra et al. (2009), sampled populations from five remote villages along the Franco-Spanish border and the Pyrenees Mountains region. The frequency of I2a-M26 ranged from 3 to 16 percent. Interestingly, ancient Y-chromosome data (see **Supplementary Table 10.10**) suggest a much broader distribution of the I2a-M26 mutation in mainland Europe at the onset of the Neolithic. This observation, along with archeological perspective previously detailed, suggest the presence of the I2a-M26 among contemporary Basques can be explained by population expansions from southern European refugia at the onset of the Holocene.

A discussion of the Basque language isolate continues in Chapter 18 and the discussion of the R1b-DF27 mutation.

Section 8. Conclusions for Haplogroup I-M170.

I-M170 mutations suggest that language has roots that extend deep into the human prehistory. Language thrives and survives because the human tribe thrives and survives. For example, the I1-M253 data, along with archeological and climate perspectives, suggest that the story of Germanic begins with the onset of the Holocene. The landscape transitioned from tundra to forest in Scandinavia, and the reindeer disappeared. People in the region adapted to climate change by harvesting marine resources. This subsistence strategy lasted for thousands of years. Prehistoric Scandinavians then adopted agriculture. Today, these cultural adaptations of the prehistoric past help to define the position attained by Germanic within the global tapestry of language variation.

Chapter 11: Haplogroup J-M304.

Section 1. The Contemporary Distribution of Haplogroup J-M304.

The website for the International Society of Genetic Genealogy regularly updates the phylogenetic tree of human Y-chromosome mutations. A comparison of the 2006 and 2020 phylogenetic trees shows that our knowledge Y-chromosome variation has achieved astonishing resolution over the last two decades. This evolving picture of human variation is made possible by researchers who have identified thousands of Y-chromosome mutations. Interestingly, the J-M304 mutation was the first polymorphic marker that was sequenced from the non-recombining region of the Y-chromosome (see Casanova et al. 1985).

At this point the reader is directed to **Supplementary Table 11.1** which provides a regional survey of populations with the J-M304 variation. As shown by the table, haplogroup J-M304 has a broad geographical distribution that extends across Eurasia and North Africa. For linguists, the J-M304 haplogroup represents a significant marker for deciphering the prehistory of several different language families: Indo-European, Afro-Asiatic, North Caucasian, Kartvelian, and Dravidian. Linguistically informative J-M304 mutations include J1-M267, J2a-M410 and J2b-M12. As shown by **Supplemental Table 11.2**, the J1-M267 mutation attains an especially high frequency among populations that speak Afro-Asiatic and North Caucasian languages. The J2a-M410 mutation is a significant marker among North Caucasian, Kartvelian, Indo-European, and Dravidian populations (see **Supplementary Table 11.3**). Finally, the J2b-M12 marker (see **Supplementary Table 11.4**) is a significant marker among Indo-European and Dravidian populations of South Asia.

Section 2. The Evolutionary History of the J-M304 Haplogroup.

At this point the reader is directed to **Supplementary Figure 1.1** from Chapter 1. As shown by the figure, haplogroups I-M170 and J-M304 have a close phylogenetic relationship. Both evolved from the IJ-M429 mutation roughly 44 thousand years ago (Poznik et al. 2016). As discussed in Chapter 10, I-M170 evolved in Europe. J-M304, on the other hand, appears to have evolved in the Caucasus or Middle East based on analysis of the ancient Y-chromosome data (see **Supplementary Tables 11.5** to **11.7**).

As noted above, linguistically informative J-M304 variants are J1-M267, J2a-M410 and J2b-M12. The reader is now directed to **Supplementary Figure 11.1** which depicts the internal phylogeny of the J-M304 main haplogroup. As shown by the figure, J-M304 bifurcates into J1-M267 and J2-M172. Poznik et al. (2016), based on analysis of contemporary data, suggest the J1-M267 mutation evolved roughly 33 thousand years ago. As shown by **Supplementary Figure 11.1**, J2-M172 splits into J2a-M410 and J2b-M12. Finocchio et al. (2018), based on analysis of contemporary data, suggest that this occurred between 12 to 16 thousand years ago. Finally, analysis of contemporary data suggests that J1-M267, J2aM410, and J2b-M12 evolved in the Caucasus or northern Zagros Mountains of Iran (see Platt et al. 2017). This conclusion is also supported by ancient Y-chromosome data (see **Supplementary Tables 11.5** and **11.6**). For example, the oldest J1-M267 remains come from Georgia in the Caucasus. Turning now to ancient J2a-M410 mutations, the oldest data come from Georgia and an individual who died about 9,700 years ago. Finally, the oldest J2b-M12 remains come from a Neolithic archeological site in the Lorestan Province of western Iran. These remains are dated to about 10 thousand years ago.

Section 3. Modeling the Origins of Afro-Asiatic.

3.1. Early Attestations of Language in Southwestern Asia.

Early attestations of Indo-European and Afro-Asiatic languages point to Southwest Asia as the putative homeland of both language families. For Indo-European, the earliest attestations are from the extinct Anatolian branch and include Hittite and Luwian. Of the two, Hittite appeared the earliest, about 3,700 years ago. Early attested Afro-Asiatic languages in Southwest Asia include Akkadian, Hebrew, Phoenician, Aramaic, and Canaanite. Among these languages, Akkadian appeared first, roughly five thousand years ago. Interestingly, around the same time early attestations of Egyptian also appeared in North Africa. Besides Afro-Asiatic and Indo-European, early attestations of language in Southwest Asia include extinct Hurro-Urartian languages such as Hattic. Finally, Sumerian and Elamite, two extinct language isolates, are also among the early attested languages of the region. For more details, see Issar and Zohar (2007: 263-272).

Figure 11.1. Manishtusu Obelisk and Akkadian Inscription. Source: Wikipedia and Mbzt.

3.2. Afro-Asiatic and Theoretical Approaches.

A synthesis of genetic, linguistic, and climatological data as presented above and in Chapter 5 suggests that proto-Afro-Asiatic arose in Southwest Asia. Semitic languages represent *in situ* diversification of Afro-Asiatic in Southwest Asia, whereas Berber, Chadic, Cushitic, Egyptian, and Omotic languages are branches of the Afro-Asiatic language family that evolved in Africa. As previously discussed in Chapter 5, the expansion of early agriculture in Southwestern Asia explains the contemporary distribution of Afro-Asiatic languages. The transition to agriculture in Southwest Asia began about 14 thousand years ago in the so-called Fertile Crescent. Over time people developed the ability to cultivate crops of legumes and grains. Additionally, the transition to agriculture involved the domestication of animals including sheep and goats. Eventually, the Southwest Asian agricultural package expanded from the Fertile Crescent into North Africa and the Arabian Peninsula.

In 1996, Semino et al. published a paper suggesting a good correlation between the distribution of J-M304 variation and the Neolithic expansion of agriculture from the Fertile Crescent of Southwest Asia. Since 1996 several other studies have reached similar conclusions (e.g., Arredi et al. 2004; Semino et al 2004; Abu-Amero 2009; Hovhannisyan et al. 2014; Singh et al. 2016), Platt et al. 2017). These observations are significant for linguists because they provide genetic support for the *early farming dispersal hypothesis*. This hypothesis, which was formulated by the archaeologist Peter Bellwood in his 2005 monograph, posits a good correlation between the expansion of early agriculture and the contemporary distribution of several language families including Afro-Asiatic and Indo-European.

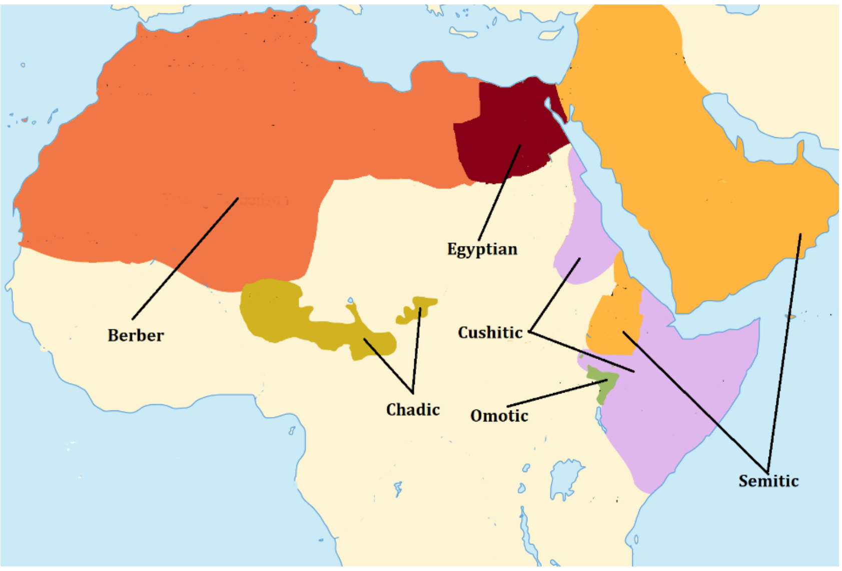
3.3. Precipitation, Language Variation, and J1-M267 Mutations.

Interpretation of the contemporary Y-chromosome data for haplogroup J-M304 is facilitated by a discussion of the climatological record in southwestern Asia. At the onset of the Holocene, roughly 15 thousand years ago, greater precipitation yielded an abundance of food resources that provided a "Garden of Eden" for the Natufian culture, the last of the hunter-gatherers in this region (see Chapter

5). Later cultural transitions, most notably the pre-pottery Neolithic B, perfected the cultivation of wild plants that the Natufians had gathered, and the domestication of animals that they once had hunted. The success of the Neolithic pre-pottery cultural tradition can be explained by adaptations to climate change. The Holocene brought greater precipitation that established a new "Mediterranean" weather pattern characterized by cool humid winters and warm dry summers. The pre-pottery cultures developed irrigation in order to bring water to the fields during the dry summer season. Furthermore, domesticated sheep and goats became an important source of food because they adapt well to arid climatic conditions. For more details, see Issar and Zohar 2007: 53-65.

Chiaroni et al. (2008) analyzed annual precipitation patterns in the Middle East during the Holocene. The researchers report a statistically significant correlation in this region between annual precipitation and the evolution of either rain fed agriculture or pastoralism. According to the study, people in the region tend to utilize rain fed (or dryland) agriculture where rainfall is abundant. Where rainfall is less than four hundred millimeters per year, people in the region tend to herd goats and sheep. Taking this a step further, the researchers suggest that the evolution of Afro-Asiatic languages reflects a specialization in semi-nomadic pastoral agriculture. According to Chiaroni et al. (2008), their model is supported by the contemporary distribution of J1-M267 and J2-M172 mutations. They suggest that J1-M267 represents the genetic signature of nomadic Afro-Asiatic-speaking pastoralists. J2-M172 mutations, on the other hand, evolved among the Indo-European-speaking farmers who cultivated crops.

Figure 11.2. Afro-Asiatic and its Primary Branches 1000-2000 Years Ago. Source: Wikipedia and Tussna.



Within the J1-M267 phylogeny, the J1a-P58 mutation has emerged as a useful marker for deciphering the expansion of early agriculture and Afro-Asiatic languages. The contemporary distribution of this mutation follows a star-like pattern where the highest frequencies are observed among the populations of Southwest Asia and smaller frequency results are reported for Eastern Europe, the Caucasus, Central Asia, and Africa (see **Supplementary Table 11.8**). From the archaeological record, the emergence of pre-pottery Neolithic B culture and its southward expansion out of Anatolia correlates well with the evolutionary history of the J1a-P58 mutation. According to Chiaroni et al. (2010), J1a-P58 evolved roughly 9,000 years ago in the Taurus and Zagros mountains of eastern Turkey. Additionally, their analysis of the data suggests that J1a-P58 mutations found in contemporary North Africa stand as a genetic relic of the expansion of southwest Asian agricultural package into the Nile River Valley roughly 6,000 years ago.

Turning now to Afro-Asiatic and the archaeological record, a recent research paper (Arbuckle and Hammer 2019) examines the development of pastoralism in the Near Ear over an eight-thousand-year period, from the Neolithic to the Bronze Age. According to the paper, pastoralism appeared in the Euphrates Valley of Turkey about 10 thousand years ago. Within 2,000 years, sheep and goat herding quickly spread across Fertile Crescent. Interestingly, it is difficult from an archaeological perspective to decouple the prehistory of herding in Southwest Asia with the evolution of irrigated crop agriculture in the region. Arbuckle and Hammer (2019) argue that pastoralism was generally tied to sedentary villages that practiced a mixed subsistence strategy of crop agricultural and the herding of goats and sheep. Later, this mixed subsistence strategy from southwestern Asia was successfully transplanted into the Nile River Valley (e.g., Haaland and Haaland 2013). Nevertheless, the archaeological record still provides evidence of nomadic pastoralism, a subsistence strategy like that of contemporary Bedouin. Data from Jebel Olaf in Saudi Arabia suggest that nomadic herders carried the Levantine agriculture package to the Arabian Peninsula as early as 8,000 years ago (see Guagnin et al. 2020).

3.4. J1-M267 and Semitic Languages.

Seventy-nine languages are classified within the Semitic branch of the Afro-Asiatic language family. Nevertheless, the sub-branches of Semitic differ among the various reference resources and language catalogues. A useful division of Semitic is provided by Hetzron (2009) and his grouping of the branch into east and west sub-branches. East Semitic consists of Akkadian, the historical language of Mesopotamia. West Semitic, on the other hand, consists of contemporary languages that include Neo-Aramaic, Hebrew, Arabic, and Amharic. In terms of geographic distribution, West Semitic is found in the Middle East, North Africa, and East Africa. The Semitic languages of East Africa, such as Amharic, stand as linguistic relics of cultural exchange between the continent and the Middle East and Africa during historical times (e.g., Gebremedhin and Mebrahtu 2020).

Based on the distribution of J1-M267 among the Arabic-speaking populations of the Middle East and North Africa, several genetic studies have taken the position that J1-M267 variation in North Africa was shaped by the historical spread of Islam (e.g., Semino et al. 2004; Capelli et al. 2006; Zalloua et al. 2008; El Sibai et al. 2009; Triki-Fendri et al. 2015). Other studies, however, suggest that J1-M267 represents much earlier agricultural expansions during the Neolithic (Arredi et al. 2004; Abu-Amero et al. 2009; Tofanelli et al. 2009b; Fadhlaoui-Zid et al. 2011a; Fadhlaoui- Zid et al. 2013). Those that favor a Neolithic expansion of J1-M267 have the historical record on their side, which presents little if any evidence to associate the spread of Islam with mass migration.

The Neolithic expansion of J1-M267 across North Africa also agrees with climatological and anthropological perspectives (e.g., Kuper and Kroepelin 2006; Linstädter 2008; Haaland and Haaland 2013). At the end of the last North African humid phase, about 7,000 years ago, sheep and goat herders expanded out of the Middle East into North Africa. It appears this expansion was driven by desiccation in the Middle East and the availability of freshwater both in the Nile Valley and along the Mediterranean coastline of northwestern Africa. Thus, Berber languages represent *in situ* diversification of Proto-Afro-Asiatic that was carried by herders into northwestern Africa during the Neolithic. Similarly, Egyptian represents *in situ* diversification of Proto-Afro-Asiatic among those that settled along in the Nile River Valley.

Section 4. Haplogroup J-M304 and Indo-European.

4.1. Climate, Topography, and Indo-European Languages.

J2-M172 mutations have become useful markers for deciphering the prehistory of Indo-European-speaking populations (see **Supplementary Tables 11.3** and **11.4**). J1-M267 mutations, on the other hand, are the genetic signature of Afro-Asiatic populations (see **Supplementary Table 11.2**). Since the putative homeland of both languages families is Southwest Asia, researchers have turned to the climatological record and other data sources from the Middle East to explain the contemporary distribution of both language families. As previously discussed, Chiaroni et al. (2008), suggest that J1-M267 and Afro-Asiatic may reflect successful adaptations to aridity. J2-M172 and Indo-European, on the other hand, may have evolved among population that utilized rain fed farming.

Another study, (Balanovsky et al. 2017b), suggest that topography shaped language variation in Southwest Asia. The researchers suggest that the distribution of languages in this region follows a lowland/upland contrast. Upland languages include Indo-European languages that were present in the Anatolian highlands during the Neolithic. Afro-Asiatic, on the other hand, represents language evolution in the lowlands of the Fertile Crescent. The subsequent westward expansion of Indo-European was then facilitated by the Anatolian plateau. The Iranian plateau, on the other hand, facilitated the eastward expansion of Indo-European into South Asia. Taking this a step further, Afro-Asiatic and Indo-European followed evolutionary trajectories dictated by topographical features that isolated early farmers from each other.

4.2. The Black Sea Deluge Hypothesis.

Analysis of the contemporary distribution of haplogroup J-M304 and G-M201 mutations strongly suggests that genes, farming, and Indo-European languages must have dispersed very rapidly from the Anatolia during the Neolithic. The dispersal pattern seems almost akin to the remains of a supernova explosion. For example, as discussed in Chapter 8, the G2a-L30 variant is found among Judeo Tats, Bagvalal, and Nogais of the Caucasus region (Karafet et al. 2016). The same mutation is also found four thousand kilometers away in Flanders (Larmuseau et al. 2014). Similarly, J2b-M241 is found both in eastern India and in Flanders (Larmuseau et al 2014, Singh et al. 2016). This distribution pattern is atypical given the phylogenetic position of both mutations and the geographic distances reported for the populations. In other words, G2a-L30 and J2b-M241 should represent *in situ* genetic diversification within a single narrowly defined geographic area.

The Black Sea deluge hypothesis provides an elegant model for explaining a rapid co-expansion of genes, farming, and language from the Black Sea. The hypothesis argues that the Black Sea was a freshwater lake at the end of the last Ice Age. According to Ryan et al. (2003), about 8,500 years ago an earthen dam collapsed at the Straights of Bosporus due to melting glacial ice and the corresponding rise of the Mediterranean Sea level. A "catastrophic flood" followed, creating the current saltwater sea that is much larger than the original freshwater lake. Based on radiocarbon and paleo-shoreline data, Turney and Brown (2007) propose this catastrophic flood occurred over a 34-year period. The researchers further suggest that around 72 thousand square kilometers of land was flooded, and this may have displaced as many as 145 thousand farmers. Moreover, Ryan and others in their 1997 discussion of the Black Sea flood suggested that this event resulted in a dispersal of farmers towards Europe, accelerating the Neolithic transition on this continent (see, also, Karachanak et al. 2013).

In 2017 several researchers revisited the Black Sea deluge hypothesis (see Yanchilina et al. 2017). They considered, once again, core sample data collected from the Black Sea. This time, however, they focused on salinification rather than an expansion of the Black Sea in terms of surface area. This follows the idea that during glacial periods the Black Sea becomes a freshwater lake because lower sea levels prevent an influx of saltwater from Mediterranean. During interglacial periods, such as the current Holocene, on the other hand, the Black Sea becomes brackish because rising sea levels reconnect this body of water with the Mediterranean Sea. Yanchilina et al. (2017) propose that the Black Sea reconnected to the Mediterranean Sea about 9,000 years ago. They also estimate that salinification of the Black Sea occurred within a forty-year period. Finally, the researchers suggest that within a nine-hundred-year period the Black Sea no longer supported freshwater fauna.

It should be emphasized that the Black Sea deluge hypothesis is controversial among researchers. For example, Yanko-Hombach et al. (2007) propose that the transition from a freshwater to saltwater lake involved a process that lasted several thousand years. Nevertheless, the data reported by Yanchilina et al. (2017) are compelling, partly because of radiocarbon results that support the rapid demise of freshwater mollusks. Additionally, the researchers emphasize general agreement among their colleagues as to when the Black Sea was re-connected to Mediterranean. Rather, disagreement involves the elevation of the water level when this occurred, a point that either supports or undermines the idea of a catastrophic flood.

One weak link with the Black Sea deluge hypothesis is the absence of robust archaeological evidence for farming settlements along the shoreline of the Black Sea at the time of the flood. The lack of archeological data might be explained by a catastrophic flood that erased archaeological evidence of Neolithic settlements. Nevertheless, ancient Y-chromosome support comes from Barcın Höyük, a Neolithic archaeological site in the Marmara Region of Northwestern Turkey. Remains found at this site are dated to around eight thousand years ago and provide Y-chromosome mutations that are linked to the Neolithic expansion of agricultural from southwestern Asia (see **Supplementary Table 11.9**).

Section 5. Haplogroup J-M304 and Language Variation in the Caucasus.

For the purposes of this discussion, the Black Sea defines the western boundary of the Caucasus region, and the eastern boundary of the region is defined by the Caspian Sea. Additionally, the Greater Caucasus Mountains define the northern boundary, and the southern boundary is defined by the Lesser Caucasus Mountains. This region includes the contemporary countries of Armenia, Georgia, and Azerbaijan as well as parts of southern Russia. Much of the archaeological data for the region comes from sites along the shores of the Caspian and Black seas and from sites along the Kura and Arax Rivers. Among archaeologists, one important question involves the arrival of the Neolithic in the region (e.g., Chataigner et al. 2014; Nishiaki et al. 2015). Some maintain that the agricultural transition was driven by the arrival of farmers from the south. Others suggest that the transition involved adoption of a new technology by people already living in the region.

Examination of the Neolithic transition in the Caucasus also elucidates the complex pattern of linguistic diversity that is found in the region. For example, an interesting study from 2019 (Beridze) suggests that the Kartvelian languages of the Caucasus represents a linguistic relic of a farming expansion from northern Mesopotamia. This conclusion is based on botanical evidence that identifies the Middle East as the source of wheat that is cultivated in Georgia. However, this conclusion is undermined by the possibility wheat was adopted by hunter-gatherers already living in the Caucasus. As such, Kartvelian is potentially a linguistic relic of hunter-gatherers of the Caucasus who maintained an indigenous language of the region during the Neolithic transition.

Turning now to the genetic data, several studies have explored Y-chromosome diversity in the Caucasus in an effort to decipher the complex linguistic diversity found in the region (e.g., Nasidze 2003; Nasidze et al. 2004; Balanovsky et al. 2011, Herrera et al. 2012; Yunusbayev et al. 2012; Hovhannisyan et al. 2014; Karafet et al. 2016; Balanovsky et al. 2017b). The frequency pattern of J1-M267 and J2-M172, as reported among the various linguistic groups of the region, is rather interesting. J1-M267 attains a heavy frequency among North Caucasian speakers, and a moderate frequency among some Armenian-speaking populations (see **Supplementary Table 11.2**). As shown by **Supplementary Table 11.3**, J2a-M410 appears cross-linguistic frequency among the population of the region. J2b-M12, on the other hand, is virtually absent (see **Supplementary Table 11.4**).

Platt et al (2017) utilize a synthesis of several different data sources to model post-glacial expansions in Southwest Asia. The researchers identify the Caucasus as a refugium during the Last Ice Age. The study further suggests that around the time of the Last Glacial Maximum, haplogroup J-M304 evolved and diversified in this region. This conclusion is supported by ancient DNA data that places the oldest J1-M267 and J2a-M410 remains in Georgia (see **Supplementary Tables 11.5** and **11.6**). Expansion of haplogroup J-M304 out of the Caucasus appears to have occurred at the beginning of the Holocene. According to the climate data, this transition brought more precipitation to the region (see Connor and Kvavadze 2009). Improved climatic conditions brought more abundant food resources along the Kura and Axas Rivers, and this may have produced population pressure that drove J-M304 mutations southwards during the Mesolithic. This suggests that J-M304 mutations represent an indigenous genetic component among the populations of the Caucasus.

As discussed previously in Chapter 8, G-M201 mutations represent potential Neolithic migrations into the Caucasus. Support for this position comes from the archaeological record which dates arrival of agriculture in the Caucasus to about 8 thousand years ago (e.g., Nishiaki et al. 2015). Additionally, the oldest G-M201 remains, which date to about 10 thousand years ago, were found at the Boncuklu in south-central Turkey (Kilinc et al. 2016). This archaeological site is located near Çatalhöyü, one of the earliest farming settlements of the Middle East. Vastly enhanced reproductive associated with early agriculture may have produced population pressure that drove G-M201 mutations from the Fertile Crescent towards the Black Sea. Later, a catastrophe flood produced an expansion of G-M201 mutations into Europe, the Caucasus, and South Asia.

Turning now to the linguistic perspective, language contact theory and the associated phenomena of language maintenance and language shift represent crucial concepts for explaining language variation in the Caucasus. The haplogroup G-M201 and J-M304 data, as discussed above, offer compelling arguments that link North Caucasian and Kartvelian with language maintenance. Similarly, the same data link the Turkic-speaking populations with a language shift model.

Language diversity in the Caucasus also includes Armenian, a branch of the Indo-European language family. Among Armenian-speaking population, haplogroup J-M304 attains a moderate frequency. Unlike North Caucasian, Kartvelian, and Turkic, however, the genetic data cannot support a language contact model to explain the presence of the Armenians in the Caucasus. Traditionally, the origins of Armenians are linked to the kingdom of Urartu that arose about 3,000 years ago near Lake Van in the Anatolia region of Turkey. As such, migration and displacement during historical times may explain why the linguistic tapestry of the Caucasus now includes the Armenian language. Moreover, deciphering the ethnogenesis of the Armenian people is extremely complicated because this question has often been influenced by nationalism rather than historical accuracy (Petrosayan 2007).

Section 6. J-M304, Iranian, Indo-Aryan, and Dravidian.

For the purposes of this discussion, South Asia consists of contemporary Pakistan and India. J1-M267 is not a significant mutation among the populations of this region (see **Supplementary Table 11.2**). On the other hand, J2a-M410 and J2b-M12 represents especially informative mutations for deciphering prehistory of Indo-Aryan and Dravidian languages (see **Supplementary Table 11.3** and **11.4**). Turning now to the archeological record, as detailed in Chapters 8 and 9, the Neolithic transition in South Asia has a Southwest Asian component. The Mehrgarh civilization settled in the Balochistan region of Pakistan about 9,000 years ago. Some of the hunter-gatherers living in the region shifted to an Indo-European language and became farmers. However, Dravidian stands as a linguistic relic of hunter-gatherers who became farmers while maintaining their language. Dravidian and Indo-European-speaking farmers then expanded into the Indus River Valley of western India. Dravidian-speaking farmers eventually migrated southwards and by around five thousand years ago they arrived in southern India and Sri Lanka. Indo-European-speaking farmers, on the other hand, migrated eastwards from the Indus Valley and by around five thousand years ago, they had settled in the Ganges River Valley.

The genetic and archeological perspectives, as presented above, link Indo-Aryan languages with diversification of Indo-Iranian in the general vicinity of the Indus River Valley. Iranian languages, on the other hand, represent diversification of Indo-Iranian that occurred somewhere west of the Indus River, probably in Iran. Genetic data support for this model of linguistic variation in South Asia is provided by Singh et al. (2016). Based on their analysis of contemporary data, the study suggests that J2-M172 mutations not only stand as the linguistic relic of a co-expansion of Dravidian and early agriculture, but also Indo-Aryan and early agriculture. Salient points from the study include:

1. J2-M172 mutations attain an overall frequency of around 12 percent in India.

2. Neolithic migrations from Southwest Asia brought J2a-M410 and J2b-M12 to South Asia.

3. The J2b-M241 mutation represents most of the J2b-M12 variation among Indians.

4. J2a-PF5197 represents most of the J2a-M410 variation.

It should be emphasized that J2b-M241 mutations extend across a vast geographical expanse from Europe to India (see **Supplementary Table 11.10**). A similar pattern is observed for J2a-M410 (see **Supplementary Table 11.3**). This spatial pattern of J2-M172 variation supports, in turn, the idea that about 9,000 years ago a catastrophic event occurred in the vicinity of the Black Sea. Some fled to South Asia, and some fled to Europe. Moreover, this spatial distribution of Y-chromosome mutations matches closely that of Indo-European languages.

Section 7. Haplogroup J-M304 and Iranian Languages.

In the Middle East, Persian and Kurdish are examples of Iranian languages, a sub-branch of the Indo-Iranian branch of the Indo-European language family. Among speakers of Iranian languages the frequency of J1-M267 and J2b-M12 is generally low, whereas the frequency of J2a-M410 is moderate (cf. **Supplementary Tables 11.2**, **11.3**, and **11.4**). As noted previously in Chapter 8 and the discussion of Martirosyan (2013), linguistic data and G-M201 mutations support the idea that Armenian, Greek, and Indo-Iranian may have been part of a dialect group at the time of the Indo-European dispersal. Interestingly, the J2a-M410 mutation also attains a significant frequency among populations that speak Greek and Armenian (see **Supplementary Table 11.3**). As such, J2a-M410 data provide additional support for the dialect group proposed by Martirosyan (2013).

Section 8. Linguistic and J-M304 Variation in Central Asia.

For the purposes of this discussion, Central Asia is defined as Afghanistan, Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan. The Central Asian Neolithic was previously discussed in Chapters 7, 8, and 9. From an anthropological perspective, the Neolithic transition in Central Asia has an indigenous component, the domestication of the horse, which occurred in north-central Kazakhstan about 5,500 years ago. The agricultural transition in this region also has a component from Southwest Asian, crops and herd animals that first appeared in Turkmenistan about 8,000 years ago.

Frachetti (2012) and Spengler et al. (2014) examine the agricultural transition in Central Asia. According to the researchers, the arrival of the Southwest Asian agricultural package in the region was facilitated by the Inner Asian Mountain Corridor, an ancient exchange route that extends from the Hindu Kush Mountains of Pakistan to the Altai Mountains of Siberia. This route is particularly significant as it facilitates travel into East Asia bypassing especially difficult terrain such as the Pamir and Tian Shan mountains ranges.

Figure 11.3. Inner Asian Mountain Corridor. Data Source: Spengler et al. (2014).

Since this route facilitated the expansion of Southwest Asian Neolithic into Central Asia, an extension of this idea would suggest that another Southwest Asian cultural commodity also flowed into Central Asia via the Inner Asian Mountain Corridor – Indo-European languages. Contemporary linguistic relics of this expansion may include Pashtun and Tajik. Genetic relics of this expansion appear to be the haplogroup G-M201, H-M2713, and J-M304 mutations that are occasionally observed among the populations of Central and East Asia.

Section 9. J-M172 and Tocharian.

Tocharian represents an extinct branch of the Indo-European language family. It was spoken until about a thousand years ago in the Tarim Basin of the Xinjiang region of eastern China. Previously, in Chapter 9, it was reported that researchers found the H1a-M69 mutation among some populations of this region. Similarly, in this region, researchers have detected the J2-M172 mutation among Turkic-speaking Uzbeks and Uyghurs as well as Iranian-speaking Tajiks (see **Supplementary Table 11.11**). Unfortunately, the published data for these populations has neglected, for the most part, to sequence J2-M172 for more informative downstream variants. Nevertheless, the available J2-M172 and H1a-M69 data strongly suggest prehistoric geneflow between southwestern Asia and East Asia via the Inner Asian Mountain Corridor. Taking this a step further, these migrations may well have carried Indo-European languages into the Tarim Basin, and this would explain the mysterious origins of Tocharian.

Section 10. J-M172, Albanian, Greek, and Italic.

As previously detailed in Chapter 8, the Neolithic transition in Europe followed two different trajectories. The Cardial Ware tradition began roughly 10 thousand years with the expansion of agriculture from Southwest Asia onto the island of Cyprus. From this location, agriculture and the Cardial Ware tradition expanded along the Mediterranean coastline of Europe and eventually arrived on the Iberian Peninsula roughly 7,500 years ago. The other trajectory, the Linear Pottery tradition, began roughly 8,500 years ago with the expansion of the Southwest Asian Neolithic into the Balkan Mountains of southeastern Europe. This expansion eventually terminated on the coastal plain of Europe roughly 7,500 years ago.

For the purposes of this discussion, Albania, mainland Greece, and mainland Italy define the Mediterranean region of Europe. Additionally, the region consists of numerous islands including Cyprus, Crete, Sicily, Sardinia, and Corsica. From a Y-chromosome perspective, the prehistory of the

Greek, Albanian, and Italic involves *in situ* diversification of Neolithic farming languages from Southwest Asia. Support for this conclusion comes from analysis of contemporary J2-M172 data. Within the region, J2-M172 mutations attain moderate frequencies among the Albanian, Greek, and Italic-speaking populations (see **Supplementary Tables 11.3** and **11.4**). According to researchers (Semino et al. 2004; Di Giacomo et al. 2004), these mutations represent a genetic relic of the Neolithic transformation in Mediterranean Europe. Ancient Y-chromosome data and the presence of J2-M172 mutations at Neolithic sites in the Middle East and Europe also support this conclusion (see **Supplementary Table 11.6**).

From an archeological perspective, the arrival of Indo-European-speaking farmers in Mediterranean Europe was carried by the Cardial Ware culture (see Chapter 8). According to archeological record (Forenbaher and Miracle 2005), these farmers appeared in eastern Adriatic roughly 8,200 years ago. The expansion from Greece to Italy may have involved a pioneering phase followed by a colonizing phase. New farming territory was initially scouted by the "pioneers." They surveyed the land and contacted hunter-gatherers living in the area. This advanced party then departed leaving behind domesticated animals for the future. Later, colonizers returned and constructed permanent farming settlements.

Section 11. The Neolithic Transition and Language Shift.

In Scandinavia, the J-M304 mutation is found in about 4 percent of Swedes (Karlsson et al. 2006). Further north in Finland, J-M304 was not detected in a study that sequenced over five hundred samples (Lappalainen et al. 2008). These data for Scandinavia follows a pattern throughout Europe whereby the frequency of J2-M172 steadily decreases from southeastern Europe. Several studies provide an explanation (e.g., Capelli et al. 2007; King et al. 2008; Battaglia et al. 2009; and Karachanak et al. 2013). They suggest that the Neolithic spread of farming from Southwest Asia to Europe was carried by farmers with the J2-M172 and G-M201 mutations. When they expanded into Europe, the continent was inhabited by hunter-gatherers. The gene pool of the hunter-gatherers included E1b-V13 (see Chapter 5), variants of the R-M207 main haplogroup Chapter 18), and variants of the I-M170 haplogroup (Chapter 10). In some cases, new populations were formed through admixture between farmers and hunter-gatherer. Alternatively, hunter-gatherers simply adopted farming. Later, agriculture facilitated rapid population growth that ultimately transformed the genetic map of Europe. This explains the Y-chromosome mutations of Neolithic, Mesolithic, and Paleolithic origin that are observed among the populations of the continent. Although the Neolithic farming genes of Southwest Asia eventually disappeared in Western European and Scandinavia, the Indo-European languages carried by the first farmers thrived and survived. In other words, Albanian, Greek, Italic, Slavic, Celtic, and Germanic are the linguistic relics of the Neolithic transformation in Europe.

The *demic diffusion* model often appears in genetic studies that discuss Neolithic transformation in Europe (e.g., Balaresque et al. 2010). Development of this model stemmed from interdisciplinary collaboration between the archaeologist Albert Ammerman and the geneticist Luca Cavalli-Sforza (e.g., Ammerman and Cavalli-Sforza 1984). Both researchers proposed that the expansion of the Southwest Asian agricultural package involved a migration of a small number of farmers into previously *uninhabited* areas. A population explosion followed their arrival because agriculture is a subsistence strategy that potentially supports high population density within a given region.

Some researchers have failed to understand that the demic diffusion model requires an agricultural expansion into an *uninhabited* area. Indeed, the demic diffusion model may explain the Neolithic transition in some regions of the world such as the Austronesian expansion into the uninhabited islands of Oceania. However, in other regions, especially Europe and South Asia, the same transition is better modeled by admixture between farmers and hunter-gatherers, or alternatively, acculturation. For linguists, the salient point here that is language shift among hunter-gatherers stands as a potential "byproduct" of the Neolithic transition in many regions of the world.

Section 12. Distribution of Ancient G-M201 and J-M304 Mutations.

According to the ancient Y-chromosome data, haplogroup G-M201 mutations surface far more often among Neolithic remains in Europe than haplogroup J-M304 mutations. The reader is reminded that ancient DNA data *confirm* rather than *exclude* mutations at a specific place and time in the prehistoric past. The genetics of the Neolithic transition in Europe is complex. The frequency of some mutations exploded while other mutations disappeared, such a I2c-L596, C1a-V20, and H2-P96.

Section 13. Conclusions for Haplogroup J-M304.

J1-M267 and J2-M172 probably evolved in the Caucasus roughly 30 thousand years ago and expanded into Anatolia and the Fertile Crescent at the beginning of the Holocene. Within the Caucasus, both mutations suggest that North Caucasian and Kartvelian languages represent an indigenous linguistic component of the region. Elsewhere, J1-M267 is linked to Afro-Asiatic languages, and J2-M172 is linked to the Indo-European family. During the Neolithic, Afro-Asiatic languages, J1-M267 mutations, and agriculture co-expanded out of the Levant into North Africa and the Arabian Peninsula. This expansion is linked to human cultural adaptations that enabled people to thrive arid condition that are characteristic of both regions. J2-M172, on the other hand, correlates well with Neolithic farmers who expanded away from the Black Sea about 9,000 years ago following a catastrophic flood. This expansion carried Indo-European languages into Europe, Iran, South Asia, and Central Asia.

Chapter 12: Haplogroups L-M20 and T-M184.

Section 1. L-M20 and T-M184 within the Y-Chromosome Phylogeny.

At this point the reader is invited to review **Supplementary Figure 1.1** from the first chapter. As shown by the figure, the LR-M9 mutation bifurcates into LT-L298 and KR-M526. Haplogroups L-M20 and T-M184 then diverge from LT-L298.

Historically, researchers have experienced difficulty in resolving the position occupied by L-M20 and T-M70 within the overall Y-chromosome phylogeny. As previously detailed in Chapter 1, the first Y-chromosome mutation was identified in 1985. By 2002 advances in sequencing technology allowed researchers to identify over two hundred Y chromosome markers. At this point, however, geneticists were utilizing at least seven different nomenclature systems to label these mutations. This, of course, hindered the potential of the Y-chromosome as a research tool and standardization was clearly needed. That year the Y-Chromosome Commission (YCC 2002) issued what is still the standard nomenclature for Y-chromosome haplogroups.

In the YCC 2002 report, L-M20 and K2-M70 both appeared downrange from K-M9. Karafet et al. (2008) then re-labeled K2-M70 as T-M70 and placed this mutation along with L-M20 downstream from K-M9. Chiaroni et al. (2009) later identified the M526 mutation as a downstream variant of M9. Shortly thereafter, Mendez et al (2011) reported the discovery of the M184 mutation, which then became haplogroup T-M184. The M70 mutation, in turn, became T1-M70. The same study also identified LT-L298 as a sister clade of M526. Finally, the study reported that LT-L298 unites T-M184 and L-M20. In 2012, the International Society of Genetic Genealogy (ISOGG) repositioned the M70 mutation within the Y-chromosome phylogeny, and T1-M70 became T1a-M70.

The discussion from the previous paragraph was provided to emphasize two important points. First, paragroups LR-M9 and its downstream mutations have been difficult to position within the Y-chromosome phylogeny. This topic will resurface in Chapter 13 and the discussion of KR-M526 paragroup. Secondly, researchers should know that data for haplogroup T-M184 must be gleaned from pre-2011 studies that report data for K2-M70 and T-M70.

Section 2. Contemporary Distribution of Haplogroups L-M20 and T-M184.

The reader is asked to note that haplogroups L-M20 and T-M184 occupy a rather small corner of the Y-chromosome map. As shown by **Supplementary Tables 12.1** and **12.2**, L-M20 and T-M184 mutations surface in Europe, Southwest Asia, the Caucasus, and South Asia. However, L-M20 is virtually absent in North and Sub-Saharan Africa, whereas T-M184 has been detected in both regions. Similarly, T-M184 is virtually absent in Central and East Asia, whereas L-M20 has been detected in both regions. Thus, the data suggest that L-M20 and T-M184 evolved in a single region and subsequently co-expanded into adjacent regions. Interestingly, a similar pattern is observed for J1-M267 and J2-M172 (see Chapter 10).

For linguists, haplogroups L-M20 and T-M184 represent important markers for deciphering the correlation between genetic and linguistic diversity. These mutations link the evolutionary history of the Afro-Asiatic, Indo-European, and Dravidian language families with the Southwest Asian Neolithic. Similar arguments were previously made for haplogroups E-M96, G-M201 and J-M304 (see Chapters 5, 8 and 11).

Section 3. Evolutionary History of Haplogroups L-M20 and T-M184.

According to Poznik et al. (2016), haplogroups L-M20 and T-M184 evolved about 45 thousand years ago. Deciphering where this occurred has been problematic. For example, Mendez et al. (2011) suggest that T-M184 evolved in the Middle East. Lacau et al. (2012) suggest that L-M20 evolved in Pakistan.

Based on the contemporary distribution of haplogroups L-M20 and T-M184, they may have evolved in the Caucasus and expanded out the region during the Holocene with haplogroup J-M304. This conclusion is supported by a recent study (Platt et al. 2021) that analyzed L1b-M317 data gathered from the Maronite religious community of northern Lebanon. The study links this marker with a population that migrated out of the Caucasus region roughly 7.3 thousand years ago. As such, L1b-M317 represents a Neolithic variant of haplogroup L-M20. Additional support comes from Mendez et al. (2011). Dating results for the T1a-L162 and T1a-L131 suggest that both markers evolved roughly 11 to 14 thousand years ago.

Support also comes from ancient DNA. The oldest T-M184 remains come from an individual who died almost 10 thousand years ago at Ain Ghazala in Jordan (see **Supplementary Table 12.8**). This archaeological site is linked to the Pre-Pottery Neolithic B cultural tradition. Additionally, T1a-M70 remains were found at Neolithic archaeological sites in Bulgaria, Germany, and Morocco. As shown by **Supplementary Table 12.7**, the oldest L-M20 remains come from Areni cave in southern Armenia. These samples were harvested from remains that date to around 6,000 years ago.

Section 4. Afro-Asiatic, Indo-European, and Dravidian.

Haplogroups L-M20 and T-M184 are useful marker for linguistics because they help to decipher the Neolithic expansion of agriculture from Southwest Asia, which began roughly 8,500 thousand years ago. The Neolithic farmers of Southwest Asia had variants of haplogroups E-M96, G-M201, J-M304, L-M20, and T-M184. When these farmers expanded out of Anatolia, their genes and languages followed. The linguistic relics of this expansion include the Afro-Asiatic and Indo-European language families. As such, Bellwood’s *early farming dispersal hypothesis* provides a robust model of prehistoric language dispersals (see, also, previous discussions in Chapters 5, 8, and 11).

Data for T-M184 and L-M20 also help to evaluate a study published by Winters in 2010. According to the report, Dravidian languages evolved in East Africa and co-expanded with the cultivation of finger millet to India. Winters (2010) cites similarities in terminology among “Africans and Dravidians” for crops. He also supports his position by claiming that the T-M70 mutation is found among East African populations as well as Dravidian speakers of India. Potential support for the position taken by Winters (2010) comes from the archaeological record. The Neolithic in South Asia has, indeed, an East African component. Furthermore, the East African Neolithic has a South Asian component. This is explained by sea trade between Africa and India. As a result of this exchange, farmers in India began to cultivate finger millet and pulses such as cowpeas, crops that they had received from Africa. East Africans, in turn, received chickens as well as Asian crops such as bananas, yams, and taro (see Fuller 2006; Crowther et al 2017).

According to the genetic data (see **Supplementary Tables 12.1** and **12.2**), both T-M184 and L-M20 are observed among Dravidian-speaking populations. However, L-M20 appears much more frequently in these populations. L-M20, on the other hand, does not appear in Africa, whereas T-M184 occasionally surfaces in some populations of North and East Africa. However, contrary to what is asserted by Winters (2010), it seems unlikely that South Asia was the source of T-M70 variation in Africa, or that Africa was the source of the same mutation in South Asia. Again, a tremendous amount of genetic, linguistic, and archaeological data, as presented here in this chapter and previously in Chapters 5, 8, and 11, all point to Southwest Asia as the source of T-M184 variation. Furthermore, these data place the likely origins of Dravidian languages in Pakistan.

Figure 12.1. East Africa and South Asia.

Section 5. Conclusions for LT-L298.

Haplogroups L-M20 and T-M184 generally attain a low frequency among the surveyed populations. Nevertheless, their geographic distribution seems to mimic that of the J1-M267 and J2-M172 mutations (see Chapter 11). Based on the available data, L-M20 and T-M184 probably evolved in Caucasus and expanded out of the region during the Holocene. For linguists, haplogroups L-M20 and T-M184 help to decipher the early expansion of Afro-Asiatic and Indo-European languages. They confirm observations derived from haplogroups E-M96, G-M201, and J-M304.

Chapter 13: The KR-M526 Paragroup.

Section 1. Introduction.

At this point it is necessary to conduct some "house-keeping." This chapter expands upon the phylogeny of LR-M9 mutations, a topic that initially surfaced in Chapter 12. The reader is now invited to review **Supplementary Figure 13.1**. The LR-M9 mutation can be found at the top of the diagram. As shown by the figure, LT-L298 and KR-M526 are sister clades that diverge from LR-M9. Paragroup LT-L298 (see Chapter 11) bifurcates into haplogroups L-M20 and T-M184. The KR-M526 paragroup, on the other hand, becomes the ancestral mutation for haplogroups M-P256 and S-B254 (see Chapter 14) as well as haplogroups N-M231, O-M175, Q-M242, and R-M207 (see Chapters 15-18).

Section 2. Nomenclature and M9 Mutations.

As discussed previously in the first chapter, the Y-Chromosome Commission standardized the nomenclature for Y-chromosome mutations in 2002. The term "paragroup" describes an intermediate mutation between Y-chromosome Adam and the haplogroups. Among these paragroups is the LR-M9 mutation. As such, it should be noted that the designation of M9 as a higher order paragroup represents a point of disagreement with Karafet et al. (2015). These researchers took the position that M9 represents a haplogroup which they label K-M9. They then argue that K-M9 has two main subclades, K1-L298 and K2-M526. Extending their argument further, the M20 and M184 mutations become subclades of K1-L298. The P256, B254, M231, M175, M242, and M207 mutations become subclades of K2-M526. Thus, for example, the R-M207 haplogroup (from Y-Chromosome Commission 2002) becomes K2b2a2-M207 (for additional information, the reader is directed to **Supplementary Figure 13.2**).

The revision proposed by Karafet et al. (2015) has been partially adopted by the International Society of Genetic Genealogy (ISOGG). The organization utilizes both the Karafet et al. (2015) nomenclature and the Y-Chromosome Commission 2002 nomenclature. However, the revision of the 2002 standard nomenclature should be rejected because it erases informative phylogenetic relationships that have distinct evolutionary histories as well as distinct patterns of geographic distribution. The Y-Chromosome Commission envisioned the concept of haplogroups as unique segments of human Y-chromosome diversity. The K-M9 haplogroup proposed by Karafet et al. (2015) clearly deviates from the standard nomenclature system envisioned by the 2002 reform. In other words, the K-M9 haplogroup proposed by Karafet et al (2015) is non-standard as it encompasses far too much of the global human Y-chromosome variation. Accordingly, in conformity with the intent of the 2002 standard nomenclature, M9 is properly classified as paragroup LR-M9, and the P256, B254, M231, M175, M242, and M207 mutations are haplogroups M-P256, S-B254, N-M231, O-M175, Q-M242, and R-M207.

Section 3. The Evolutionary History of the KR-M526 Paragroup.

The reader is now directed to **Supplementary Figure 1.1** from the first chapter and **Supplementary Figure 13.1** from this present chapter (Chapter 13). As shown by the figures, the LT-L298 and KR-M526 mutations are "sister" clades downstream from the LR-M9 paragroup. Determining where and when KR-M526 evolved is difficult. Based on the current geographic distribution of LT-L298 "sister" clade (see Chapter 12), KR-M526 probably evolved in the Middle East or the Caucasus. This occurred at least 54 thousand years ago based on dating results obtained by Bergstrom et al. (2016) with the aid of sophisticated whole genome sequencing.

This dating estimate from Bergstrom et al. (2016), along with the contemporary distribution of haplogroups M-P256, S-B254, N-M231, O-M175, Q-M242, and R-M207 (see Chapters 14 to 18), suggests that KR-M526 stands as a genetic relic of human expansions from the Levant during Marine Isotope Stage 3. This, of course raises another question, whether KR-M526 expanded eastwards along the southern or northern route. Based on data from this chapter (Chapter 13), along with that from Chapters 14 to 18, KR-M526 probably expanded along both routes. Haplogroups M-P256 and S-B254 are genetic relics of the southern expansion. Haplogroups N-M231, O-M175, Q-M242, and R-M207 are genetic relics of expansions along the northern route.

Section 4. "Uncharted" KR-M526 Mutations.

A considerable amount of data for populations in Island Southeast Asia, Oceania, and Australia consists of the frequency results for KR-M526*, SM-P399*, and P-P295* (see **Supplementary Table 13.1**, **13.2**, and **13.3**). The reader will notice the use of an asterisk. This follows the 2002 Y-Chromosome Commission standard to denote unresolved mutations. Taking this a step further, part of the Y-chromosome variation for these regions remains poorly resolved. To better understand the concept of resolution, phylogenetic relationships within the KR-M526 paragroup should be considered analogous to a map. Haplogroups M-P256, S-B254, N-M231, O-M175, Q-M242, and R-M207 are clearly on the "map." However, KR-M526*, SM-P399*, and P-P295* represent unknown mutations that lie in "uncharted waters."

Section 5. The Linguistic Distribution of KR-M526*, SM-P399*, and P-P295*.

As discussed above, among the populations of Australia, eastern Indonesia, and Papua New Guinea, a significant number of men have a Y-chromosome mutation that has not been identified. Linguistically, KR-M526* appears in published data for Australian aboriginals. KR-M526*, SM-P399*, and P-P295* appears in published data for Austronesian and Papuan-speaking populations in Island Southeast Asia and Oceania. The amount of unresolved data for these populations is unusual when compared to other locations in the world. Moreover, the unresolved data is disturbing as Australian, Papuan, and Austronesian languages occupy a large corner of the linguistic tapestry of language variation. The Australian family consists of 379 languages. Over 1,200 languages are classified as Austronesian (*Ethnologue* 2018). Papuan is a macro-language family that consists of thirty-six language families and over eight hundred languages (see **Supplementary Table 13.4**).

Section 6. KR-M526* among Australian Aboriginals.

The reader is directed to **Supplementary Table 13.1**. Unresolved KR-M526 mutations (or KR-M526*) represents about 12 percent of the indigenous Y-chromosome variation among Australian aboriginals. Enormous time depth and social factors might explain this observation. Turning now to the question of time depth, Bergstrom et al. (2016) suggest that the genetic trail leading to the founding populations of Australia may well have faded over the last 50 thousand years. This conclusion stems from whole genome sequencing of 13 samples that were collected from Australian aboriginals. These data were then compared with that from other populations. According to the researchers, KR-M526* represent ancient lineages brought to Island Southeast Asia and Australia by those who settled in both regions during Marine Isotope Stage 3. This conclusion is consistent with the archeological record (see Chapter 4). Moreover, the conclusion agrees with data for haplogroup C1-F3393. As previously detailed in Chapter 6, the founding populations of Island Southeast Asia and Australia had the C1b2-B477 mutation. C1b2a-M38 represents *in situ* evolution of C1b-B477 in Island Southeast Asia, and C1b2b-M347 represents *in situ* evolution of C1b-B477 in Australia.

Figure 13.1. Uluru (Ayers Rock), Northern Territory, Australia. Source: Wikipedia and Stuart Edwards.

Concerning social factors that may have reduced male genetic variation in Australia, Nagle et al (2016a) screened police databases for men who identified their ethnicity as Australian aboriginal. From a total of about 1,200, roughly half of them had a Y-chromosome haplogroup that is not indigenous to Australia. This suggests that over the last two hundred years, substantial admixture has occurred between men of European descent and Australian aboriginal women. As a result, a population bottleneck effect may have reduced male genetic variation among aboriginal males. According to Nagle et al. (2016a), this may hinder better resolution of KR-M526. Interestingly, a similar bottleneck effect may have also occurred among the Native Americans (see Chapter 17 for more details).

Previously, in Chapter 6, C1-M347 was identified an Australian-specific mutation which supports the position that language evolved at least 100 thousand years ago. Like C1-M347, KR-M526* also represents part of the indigenous genome among Australian aboriginals. Furthermore, like C1-M357, analysis of contemporary KR-M526* mutations among Australian aboriginals suggests that modern humans entered Australia roughly 50 thousand years ago, and their descents remained isolated on the continent until the arrival of Europeans about two hundred years ago (Bergstrom et al. 2016; Nagle et al. 2016a). Given that Australian aboriginals are strongly linked to the out-of-Africa exodus, the KR-M526* data for this populations support the position that humans acquired language before leaving Africa (again, about 100 thousand years ago). The less plausible alternative scenario would posit that language evolved independently in several regions of the world.

Section 7. Unresolved KR-M526 mutations among Papuans and Austronesians.

Insight into the complex correlation between linguistic and genetic variation in the highlands of Papua New Guinea is illustrated by a paper (Li et al.) published in 2020. Over the last 50 thousand years marriage customs and geographical barriers has generally constrained genetic variation and fostered linguistic diversity. For example, Kuru, a disease of the nervous system, reached epidemic proportions in the highlands in 1957. Today, Kuru has disappeared as the result of a government enforced ban on endocannibalism that began in the 1960s. The disease is caused by eating human brains, which was part of the mortuary practices at the time. Women performed the ritual, and for this reason they and not the men contracted Kuru. The data gathered by Li et al. (2020) suggests that the Kuru epidemic increased maternally inherited mitochondrial DNA diversity in the highlands. Because of the high rate of female mortality, men were forced to procure brides from outside their villages.

As shown by Supplemental Table 13.1, frequency data for KR-M526* is reported for Papuan and Austronesian-speaking populations in Island Southeast Asia and Oceania. Among the Papuans, this observation can be explained by enormous time depth. Among the Austronesians, the observed frequency of KR-M526* is explained by admixture with Papuans that began three thousand years ago.

Like Australian aboriginals, the genetic "trail" that links Papuans to Y-chromosome Adam has faded. Social factors may also play a role. Kayser et al. (2003) suggest that a long-standing tradition of warfare between the various tribes of New Guinea has reduced genetic variation among men in this area of the world. This loss of genetic diversity is akin to what known as a bottleneck effect (see Chapter 1 for more details). According to the same study, another factor that may have reduced male genetic variation in Island Southeast Asia is the prevalence of polygyny, the practice of having more than one wife.

Section 8. The SM-P399* Mutation.

The SM-P399 mutation stands as a paragroup downstream from KR-M526 (see **Supplementary Figure 13.1**). Additionally, the SM-P399 paragroup represents the ancestral mutation for haplogroups M-P256 and S-B254 (see Chapter 14 for more details). Two studies (Tumonggor et al. 2014; and Karafet et al. 2015) report SM-P399* data for Austronesian and Papuan-speaking populations in Indonesia (see **Supplementary Table 13.2**).

The discovery of Tianyuan Man was previously discussed in Chapter 4. These remains come from an individual who died roughly 40 thousand years ago near Beijing, China. Professor David Reich at Harvard University periodically releases a dataset that reports ancient Y chromosome data. The release from March 1, 2020 (V42.4) reported that Tianyuan belong to paragroup K2b-P331, the ancestral clade for the SM-P399 and P-P295 paragroups (see, also, **Supplementary Figure 13.1**). As previously detailed in Chapters 4 and 6, human expansions into East Asia during Marine Isotope Stage 3 followed two different routes, a southern route, and a northern route. Tianyuan Man certainly provides archeological support for the co-expansion of hunter-gatherers and KR-M526 mutations along the northern route. However, this would assume that his mutation later evolved into P-P295 rather than SM-399.

Section 9. The P-P295* Mutation.

The P-P295 paragroup has two known subclades, P1-M45 and P2-F20148. P1-M45 is the ancestral mutation for the Q-M242 and R-M207 haplogroups (see **Supplementary Figure 13.1**). According to the International Society for Genetic Genealogy, previously reported P-P295 data for Eurasian population failed to sequence for haplogroup Q-M242. The organization also reports that the P2-F20148 mutation has been observed among the Aeta, a Negrito population of the Philippines (see ISOGG 2019-2020). As such, the P-P295 data reported for Indonesia (see **Supplementary Table 13.3**) poses an interesting question, whether resequencing would place these populations within the P2-F20148 clade.

Section 10. Conclusions for the KR-M526 Paragroup.

Researchers should identify the M9 mutation as a paragroup and not as a haplogroup. The KR-M526 paragroup, a downstream variant of LR-M9, represents an important genetic relic of human expansions during Marine Isotope Stage 3. Several haplogroups eventually evolved from KR-M526. These haplogroups represent well-resolved sections of the KR-M526 phylogenetic map. Nevertheless, the literature presents unresolved mutations within the paragroup that still need to be charted on the Y-chromosome map. Better resolution of these mutations may provide greater insight into the evolution of Australian and Papuan languages. Interestingly, the significant presence of unresolved mutations for these populations underscores the idea that Australian and Papuan languages have long roots that extend deep into the prehistory.

Chapter 14: Haplogroups M-P256 and S-B254.

Section 1. Contemporary Distribution of Haplogroups M-P256 and S-B254.

This chapter presents two phylogenetically close variants of paragroup KR-M526, haplogroups M-P256 and S-B254. In order to discuss the contemporary distribution of both haplogroups, it is necessary, at this point, to review geographical terminology that was initially presented in Chapter 6 and the discussion of haplogroup C1-F3393 mutations. Island Southeast Asia includes the Philippines, Indonesia, East Timor, and Papua New Guinea. Oceania, on the other hand, consists of a broad expanse of islands in the Pacific Ocean that runs eastwards from the Solomon Islands to Rapa Nui, and southwards from the Hawaiian Islands to New Zealand. Additionally, the so-called "Wallace Line" is a useful geographical term that appeared in Chapter 6. The Wallace Line originally surfaced in the field of botany to delineate different ecozones. Over time, it evolved into a convenient political boundary that separates western Indonesia from eastern Indonesia, and a regional boundary that separates East Asia from Island Southeast Asia. Interestingly, the Wallace Line also stands as a genetic divide. Haplogroups M-P256 and S-B254 are found east of Wallace Line and are virtually absent west of the boundary (Karafet et al. 2010). As discussed in Chapter 6, a similar pattern is also observed for the C1b-M38 mutation.

The contemporary distribution of haplogroups M-P256 and S-B254 is detailed in **Supplementary Tables 14.1** and **14.2**. The reader is also directed to **Supplementary Figure 14.1** which details the internal phylogeny of both haplogroups. Geographically, M-P256 and its downstream variants represent significant mutations for deciphering the genetic history of Island Southeast Asia. S-B254 and its downstream variants represent significant mutations not only for the populations of Island Southeast Asia, but also for the Australian aboriginals. Linguistically, M-P256 and S-B254 represent significant mutations among the Papuan and Austronesian-speaking populations of Island Southeast Asia. Haplogroup S-B254 also presents data for deciphering the prehistory of the Australian language family.

Section 2. Evolutionary History of Haplogroups M-P256 and S-B254.

The reader is now invited to review **Supplementary Figure 1.1** from the first chapter. The SM-P399 mutation is a downstream variant of the KR-M526 paragroup. Haplogroups M-P256 and S-B254 then diverge from SM-P399. The SM-P399 paragroup and its downstream mutations are the genetic relics of the arrival of modern human in Island Southeast Asia. Dating estimates from Karmin et al. (2015) support this position. According to the report, haplogroups M-P256 and S-B254 evolved roughly 50 thousand years.

Section 3. Trans-New Guinea and the Early Farming Dispersal Hypothesis.

The so-called Papuan macro-language family stands as a linguistic relic of human migrations into Sunda and Sahul roughly 50 thousand years ago (see Chapters 6 and 13). As detailed in **Supplementary Table 13.4** from Chapter 13, the Papuan macro-family consists of over 800 languages that are classified into one of 36 language families. Among these language families, Trans-New Guinea occupies a unique position within the macro-language group. It is the largest Papuan family, both in terms of number of speakers (about 3.5 million) and number of languages (almost 500). Additionally, the distribution of Trans-New Guinea also extends across a far greater range than the other Papuan families, from the Wallace Line to Oceania, whereas most of the other Papuan languages are restricted to a much smaller geographical area.

A discussion of the evolution and expansion of Trans-New Guinea languages necessitates a brief discussion of New Guinea in terms of its political divisions, geographical location, and unique topography. New Guinea is the second largest island of the world. The eastern half of the island belongs to the country of Papua New Guinea and the western half to Indonesia. The lowlands define the low-lying coastal areas of the island, whereas the inland region is called the highlands. This contrast in altitude is the result of colliding tectonic plates which have pushed the center of the island upwards, forming a two-thousand-kilometer-long "spine" running east to west across the islands. Here the altitude eventually climbs to 4,000 meters above sea level. Extending from both sides of the mountain chain are numerous valleys inhabited by people (for more details, see Allen 1992).

Figure 14.1. Relief Map Showing the Central Highlands of New Guinea.

Topography also sets the stage for explaining the position that Trans-New Guinean languages occupy within the linguistic tapestry of Island Southeast Asia. Specifically, researchers have identified the central highlands of Papua New Guinea as the putative homeland of the Trans-New Guinea language family (i.e., Bellwood 2005: 142-145; Pawley 2005; Schapper 2017). An examination of the prehistory of Trans-New Guinean languages raises an interesting question, why the populations of New Guinea migrated from the lowlands into the highlands at the end of the Pleistocene. One possible explanation is malaria avoidance as the prevalence of this affliction diminishes at greater altitude. Interestingly, a recent study (Gaffney et al. 2021) suggests that by the time of the arrival of the early Holocene on Papua New Guinea, humans had depleted much of the large game food resources. In order to acquire protein, humans adapted and began to harvest small game. As such, early Holocene migrations from lowlands to highlands might be linked to the hunting forays into the montane tropical forest and the harvesting of possums, fruit bats, and other similar sized animals.

The *early farming dispersal hypothesis* was formulated by the archaeologist Peter Bellwood. According to the hypothesis, the expansion of early agriculture explains the current distribution of many of the world's language families. Bellwood (2005: 142-145), suggests that this includes Trans-New Guinea languages. Denham et al. (2003) provide archaeological and archaeobotanical context for this model. They collected data near a tea plantation at the Kuk Swamp in the Waghi Valley of the central highlands, which about 1,500 meters above sea level. The study suggests that the agriculture transition began in Papua New Guinea about 10 thousand years ago with the construction of drainage ditches. The intensive cultivation of taro root and bananas later evolved by around 7,000 years ago. Both

Denham et al. (2003) and Bellwood (2005: 142-145) emphasize that climate change facilitated the development of agriculture in the Papua New Guinean highlands. The Holocene brought higher temperatures and regular rainfall. This enabled people to exploit the fertile soil that accumulated in the highland valleys during the Last Ice Age.

Schapper (2017) presents a fascinating examination of early agriculture in the Papua New Guinean highlands and the expansion of Trans-New Guinean languages. She disputes the traditional assumption that correlates this expansion with taro root cultivation, a staple crop of the region before the arrival of sweet potatoes. Schapper suggests that the expansion of the Trans-New Guinea language family correlates better with the cultivation of bananas and sugar cane. Her argument is partly based on linguistic data. According to the researcher, proto-Trans-New Guinean reconstructions for banana and sugar cane are more reliable than those for taro root. Schapper also considers the botanical evidence. According to the researcher, taro root originally came from South Asia and eventually spread eastward across New Guinea during prehistoric times. This expansion pattern runs in the opposite direction as that of Trans-New Guineas languages. Rather, the linguistic evidence points to the highlands of Papua New Guinea as the putative homeland of Trans New Guinea languages and a westward expansion of these languages to the Wallace Line. This is based on the diversity of higher order language family sub-groups that are found in the highlands. Moreover, because bananas and sugar cane are indigenous to New Guinea, a westward co-expansion of Trans-New Guinean languages, bananas, and sugar cane is easier to defend than a co-expansion of taro root and language. Additionally, bananas and sugar cane are more versatile than taro root, and as such, are better suited for supporting a rapid westward population expansion that terminated at the Wallace Line. Compared to taro root, bananas and sugar came grow at a greater range of altitudes and soil conditions. Their cultivation is also less labor intensive.

The reader is invited to review **Supplementary Tables 14.3** and **14.4** which provide frequency data for the M1a-P34 and S1a-M254 mutations. Both mutations have been identified as the genetic signature of the Trans New Guinean expansion (Mona et al. 2007; Tumonggor et al. 2014). Particularly persuasive support for this position is provided by the dating estimates for both mutations. According to Mona et al. (2007), M1a-P34 and S1a-M254 evolved roughly 7,000 years ago in Papua New Guinea. As such, the agriculture transition in the highlands of Papua New Guinea produced a genetic “scar” that is recorded on human Y-chromosome.

As noted above, Trans-New Guinean languages evolved in the highlands of New Guinea. Unfortunately, almost all the data for the populations of New Guinea come from low-lying coastal area on the island. The only Y-chromosome data reported for the highlands of Papua New Guinea come from 31 samples sequenced by Kayser et al. (2006) and little is known about the individuals who furnished the samples. These samples were initially collected from placental tissue by Stoneking et al. (1990) for an early mtDNA study. According to the 1990 study, the tissue samples came from several villages in the highlands from people who spoke “non-Austronesian.” Their “non-Austronesian” language or languages probably belong to the Trans-New Guinea language family. This is based on a comparison of language maps prepared for Papua New Guinea by *Ethnologue* (2018) with a map furnished by the 1990 study.

Another disturbing matter that needs to be brought to the reader’s attention concerns the M1a-P34 mutation. On January 9, 2017, the International Society of Genetic Genealogy (ISOGG) removed M1a-P34 from the Y-DNA haplogroup tree because the mutation fails to meet their “quality guidelines.” This development is problematic because seven different studies report P34 data for 2,496 men and now researchers lack certainty as to where the mutation is positioned within the phylogeny of haplogroup M-P256.

Finally, it should be emphasized that lack of research for Papuan and Trans New Guinean languages stands as a serious deficiency within the field of linguists. Andrew Pawley writes the following:

> There is not a single linguist whose primary research field is Papuan historical linguistics. Only a handful of linguists are active in [Trans-New Guinean] historical studies. It might be said that studies of the Trans New Guinea family are about where Indo-European studies were in the 1820s, in the days of Rask and Grimm, but with the prospect of having only a tiny fraction of the manpower that was available for the study of Indo-European (2005: 99-100).

Nevertheless, despite all these limitations, the available genetic, archaeological, botanical, and linguistic data still point to the *early farming dispersal hypothesis* as a robust model for explaining the prehistoric expansion of Trans-New Guinean languages.

Section 4. Australian Languages.

Researching the genetic history of Australian aboriginals has been problematic. Holst Pellekaan (2013) provides an explanation stating that aboriginal Australians are reluctant to participate in genetic studies due to historical mistrust between themselves and Europeans. Of course, the paucity of data for aboriginal Australians is extremely unfortunate for linguists because we lack genetic data for over three hundred languages classified within the Australian language family. Furthermore, according to Ethnologue (2018) only 185 Australian languages are still spoken and many of these living languages face an uncertain future. Finally, as noted earlier in Chapters 6 and 13, the genetic history of Australian aboriginals helps to define when language evolved as a modern human behavioral adaptation.

Most of the Y-chromosome data for Australian aboriginals come from Nagle et al. (2016a). Their study certainly represents a step in the right direction. However, the quality of their data is mediocre, a point that the study seems to concede. Good quality data focuses on populations and provides ethno-linguistic details. Nagle et al. (2016a), however, gathered most of their data from police databases from men who identify themselves as aboriginal Australians. Unfortunately, the ethnographic details are missing and as such, researchers are unable to correlate the genetic data with a specific Australian language.

As noted earlier in Chapter 6, C1b-M347 has emerged as a unique Australian-specific mutation that represents a genetic artifact of the initial human settlement of Australia 40 to 50 thousand years ago. Undefined mutations within KR-526 are also a relic of this of humans on the continent (see Chapter 13). In their 2016a study, Nagle and others identify the S1a-P308 mutation as an Australian specific genetic artifact of this event. This explains why the mutation attains a frequency of about 12 percent among the Australian aboriginals (see, also, **Supplementary Table 14.5**).

Nagle et al (2016a) assert, based on their analysis of the Y-chromosome data, that after humans had colonized Australia between 40 and 50 thousand years ago, the ancestors of contemporary aboriginals remained isolated from the rest of the world until the arrival of Europeans in the eighteenth century. It should be noted that data from mitochondrial DNA (mtDNA) also support this conclusion (Nagle et al. 2016b; Pedro et al. 2020). The significance of the mtDNA data stems from the fact that this marker is inherited maternally whereas Y-chromosome mutations are paternally inherited. Thus, researchers can eliminate the possibility of female gene flow at a later time. Taking this a step further, both mtDNA data and Y-chromosome data support the following argument: Australian languages have roots that extend deep into the prehistory of modern humans.

During the 40 to 50-thousand-year period that Australian populations were isolated from the rest of the world, very limited gene flow may have occurred via the Torres Strait, where the distance between Papua New Guinea and Australia narrows to about 150 kilometers. This is based on presence of the M1-M4 mutation that is detected in about one percent of the Australian aboriginals, as reported by Nagle et al (2016a). According to the study, the significance of M1-M4 among Australian aboriginals needs further analysis. A future investigation of this matter will require the collection of more data from Queensland, the Australian state closest to Papua New Guinea.

Section 5. Language Contact in Island Southeast Asia.

Around four thousand ago, Austronesian-speaking people migrated from Taiwan to the Philippines. From the Philippines, Austronesian eventually expanded to eastern Malaysia, Indonesia, and Papua New Guinea. From an archeological perspective, this expansion was carried by the so-called Lapita culture which is often identified by discarded fragments of a distinctive style of pottery. Before the arrival of the Austronesians, the people of Island Southeast Asia had spoken Papuan languages exclusively. Consequently, Papuan languages represent the indigenous linguist component of language variation in the Island Southeast Asia (e.g., Pawley et al. 2005). Moreover, they are potential linguistic relics of the initial human colonization of this region around 40 to 50 thousand years ago. Haplogroups M-P256 and S-B254, on the other hand, represent the genetic relics. Thus, the widespread presence of both mutations in eastern Indonesia and Papua New Guinea among contemporary Austronesian-speaking populations is significant (see **Supplementary Tables 14.1** and **14.2**). Language shift and language maintenance have clearly forged language variation in Island Southeast Asia.

The Philippines, East Timor, Indonesia, and Papua New Guinea were colonized by modern human at roughly the same time, 40 to 50 thousand years ago, during Marine Isotope Stage 3 (e.g., Delfin 2015). Additionally, these countries lie within the initial southward expansion zone of Austronesian languages that occurred much later. However, it should be noted that the indigenous Papuan languages of eastern Indonesia, East Timor, and Papua New Guinea managed to survive after the arrival of the Austronesians, about 4,000 years ago. In the Philippines, on the other hand, Austronesian completely replaced the indigenous Negrito languages.

As previously detailed in Chapter 4, the so-called Negrito populations found in various part of Asia, including the Philippines, are potential relic populations of the out-of-Africa expansion. The Jarawa and Onge, two Negrito populations of the Andaman Islands, retain a genetic signature of this migration. Among the Negrito populations of the Philippines, however, the data are inconclusive as to whether they still retain relic mutations. The best data for Filipino Negritos are reported by Delfin et al. (2011) and unfortunately the study utilized poor resolution markers, namely C-M130 and K-M9. The samples gathered by Delfin et al. (2011) should be re-sequenced for more informative markers. One compelling reason for taking this step is that Karafet et al. (2015) report data for the Aeta, one of the Negrito populations of the Philippines. According to the study, among the Aeta the S2-P378 mutation attains a frequency of 60 percent. This finding suggests that re-sequencing of the Delfin et al. (2011) samples may detect additional Negrito populations with S2-P378. This mutation potentially links the population of the Philippines with the rest of Island Southeast Asia, and this would provide valuable insight about language shift and language maintenance in the region.

Focusing on the Austronesia populations of the Philippines, it should be noted that Delfin et al (2011) provide the only source of published data for Y-chromosome mutations. Moreover, the study reported data for 210 Austronesian samples using poor resolution markers. Since 104 million people live in the Philippines (CIA World Factbook 2018), more data collection and analysis are necessary.

Section 6. Conclusions for Haplogroups M-P256 and S-B254.

Haplogroups M-P256 and S-B254 have emerged as important markers for understanding the genetic history of populations in Island Southeast Asia and Australia. For linguists, these mutations are important markers for deciphering the prehistory of Papuan, Australian and Austronesian languages. Data from both mutations suggest that language shift and language maintenance have clearly forged language variation in Island Southeast Asia. The indigenous Papuan languages of eastern Indonesia, East Timor, and Papua New Guinea managed to survive after the arrival of the Austronesians. However, Austronesian completely replaced the indigenous languages of the Philippines. Additionally, the M1a-P54 and S1a-M254 mutation are informative markers for explaining the expansion of Trans New Guinea languages. These genetic data, along with the archaeological, botanical, and linguistic evidence, suggest that this expansion conforms to the *early farming dispersal hypothesis.* Finally, the S1a-P308 mutation represents an important marker for deciphering the genetic history of aboriginal Australians. This, in turn, provides a convenient scale to assess the minimal age of human language. We witness the potential evolution of over 300 indigenous Australian languages from a language spoken by a very ancient population that is closely linked to the out-of-Africa exodus. This suggests the following: When our ancestors first left Africa, among the tools they carried for the journey was language.

Chapter 15: Haplogroup N-M231.

Section 1. The Contemporary Distribution of N-M231.

The N-M231 haplogroup and its downstream variants help to decipher genetic diversity throughout a vast area that consists of Northern Eurasia, the Baltic Region, Scandinavia, and Eastern Europe. Additionally, the same data help to resolve the prehistory of several different language groups including Baltic, East Slavic, Germanic, Tungusic, Turkic, Mongolic, Chukotka-Kamchatkan, and Eskimo-Aleut.

It should be noted that haplogroups N-M231 and O-M175 share a common ancestral mutation, the NO-M214 paragroup. Nevertheless, despite the phylogenetic closeness, both haplogroups have very different contemporary geographic distributions. Haplogroup O-M175 represents an important marker in East Asia, South Asia, and Oceania, whereas haplogroup N-M231 attains a frequency of about only about 6 percent in East Asia (Zhong et al. 2011). See, also, Chapter 15. Additionally, N-M231 is almost absent in South Asia and Oceania. Instead, haplogroup N-M231 stands as a marker of North Eurasian populations with a geographic distribution that extends over a vast territory from the Pacific Ocean in the East to the Atlantic Ocean in the West. O-M175, on the other hand, is virtually absent in Northern Eurasia.

The reader is now directed to **Supplementary Figure 15.1**, which provides a phylogenetic overview of haplogroup N-M231. The internal phylogeny of the N-M231 haplogroup has two main divisions, N1a-F1206 and N1b-F2930. The N1b-F2930 mutation represents almost all of the haplogroup N-M231 variation in East Asia (see **Supplementary Table 15.1**). N1a-F1206, on the other hand, represents haplogroup N-M231 variation in Northern Eurasia. Most of the published data for N1a-F1206 consists of N1a-P43 and N1a-M46 mutations. Both mutations are distributed over a vast geographical expanse, from the Baltic region to eastern Siberia (see **Supplementary Tables 15.2** and **15.3**).

The N1a-P43 mutation is not a particularly informative as its internal phylogeny remains poorly defined. Researchers, on the other hand, have been able to obtain far better resolution of the N1a-M46 mutation. The discovery of the mutation was reported by Zerjal et al. in 1997. In many of the early genetic studies that followed, researchers often identified the M46 as the “Tat” mutation. The term “Tat” describes an unusual category of genetic mutations that help to define this haplogroup.

The reader may want to review **Supplementary Figure 15.1** and the phylogenetic overview of haplogroup N-M231. Informative N1a-M46 include N1a-B211, N1a-Z1936, N1a-M2019, N1a-VL29, N1a-B479, N1a-F4205, and N1a-B202. The N1a-B211 mutation is found in Northern Eurasia and Eastern Europe among populations that speak Uralic or Turkic languages (see **Supplementary Table 15.4**). As shown by **Supplementary Table 15.5**, N1a-Z1936 is found in Scandinavia among Finns and Sami. It is also found among Russians, Veps and Karelians in the Baltic region, and in Northern Eurasia among Turkic-speaking Tatars and Uralic-speaking Komi. N1a-M2019 is found in Northern Eurasia among Tungusic and Turkic speaking populations (see **Supplementary Table 15.6**). Among Estonians, Latvians, and Lithuanians, N1a-VL29 attains a significant frequency (see **Supplementary Table 15.7**). The same mutation is found among the Uralic-speaking populations of Scandinavia, Eastern Europe, and Northern Eurasia. N1a-B479 appears to be the genetic signature of the Nanai, a Tungusic-speaking people in Eastern Siberia among whom the mutation attains a frequency of over 40 percent (see Illumae et al. 2016). The N1a-F4205 mutation (see **Supplementary Table 15.8**) appears to be a unique mutation of Mongolic-speaking populations in East Asia and Northern Eurasia. Finally, as shown by **Supplementary Table 14.9**, the N1a-B202 mutation is found in Eastern Siberia among the Chukchi, Koryaks, and Yupik.

Section 2. The Evolutionary History of N-M231.

2.1. Origins of Paragroup NO-M214.

At this point the reader should review **Supplementary Figure 1.1** from the first chapter. The NO-M214 mutation is a downstream variant of KR-M526. Poznik et al. (2016) suggests that NO-M214 evolved roughly 47 thousand years ago. To determine where this occurred, it is necessary to revisit the out-of-Africa model that was previously introduced in Chapter 4. *Homo sapiens* migrated from eastern Africa into the Levant about 130 to 100 thousand years ago during Marine Isotope Stage (MIS) 5, around the beginning of the last Ice Age. For a period of between 50 and 80 thousand years, human populations in the Levant expanded, contracted, and re-expanded. During Marine Isotope Stage 3, about 50 thousand years ago, a temporary amelioration of widely fluctuating Ice Age climatic conditions facilitated human migrations out of the Levant. For researchers, these migrations signal the expansion of modern human into Europe, Northern Eurasia, East Asia, and Australia.

Traditional interpretation of the archaeological, genetic, and paleo-climatological data posit human colonization of East Asia via a *single* migration route during Marine Isotope Stage 3 (e.g., Mellars 2006; Pope and Terrell 2008; Stoneking and Delfin 2010; Oppenheimer 2012; Karafet et al. 2015). According to these reports, this so-called “southern route” initially followed the coastline of southern Asia. In southeastern Asia, some proceeded northwards along the coastline of East Asia into Japan. Other traveled south into Island Southeast Asia and Australia. From a Y-chromosome perspective, genetic relics of the southern migration model are haplogroups D-M174, C1-F3393, M-P256, and S-B254. However, a *single* human expansion into East Asia via the “southern route” is inconsistent with ancient DNA data. Rather the data also support a second expansion into East Asia via the “northern route.”

As presented in **Supplemental Table 15.10**, ancient Y-chromosome support for the “northern route” hypothesis partially consists of the C1-F3393 remains found in Europe. Additionally, the data consists of the Tianyuan remains, an individual who had the ancestral marker for haplogroups Q-M242 and R-M207. The data also consist of two remains that have the NO-M214 paragroup, the ancestral mutation that unites haplogroups O-M175 and N-M231. One set of NO-M214 remains comes from the Ust’-Ishim man, an individual who died about 45 thousand years in western Siberia. The other set of remains, Oase1, comes from the Peştera cu Oase cave in Romania. He died roughly 40 thousand years ago.

As previously detailed in Chapter 6, modern humans expanded onto the Central Eurasian steppes roughly 50 thousand years ago to hunt the large herbivores that once roamed this region. This expansion into Central Eurasia consisted of men with KR-M526 mutations (see Chapter 13). A synthesis of climate, genetic and archeological perspectives suggests that somewhere on the steppes, NO-M214 evolved from KR-M526. Then another population split occurred. One group carried NO-M214 from Central Eurasia to Eastern Europe. The other group carried NO-M214 eastwards from Central Eurasia to China and Mongolia.

2.2. Diversification of the N-M231 Haplogroup.

Haplogroup N-M231 and its sister clade, O-M175, evolved from NO-M214 about 42 thousand years ago (Poznik et al. 2016). The genetic and paleoclimatological evidence (e.g., Shi et al. 2013) suggest that this occurred in China. Additional support comes from a 2015 report published by Hu et al. Haplogroup N-M231 has two main internal divisions, N1a-F1206 and N1b-F2930. According to Hu et al. (2015), both mutations evolved roughly 16 thousand years ago in China. As previously mentioned above, N1b-F2930 eventually became the dominate haplogroup N-M231 variant in East Asia and N1a-F1206 represents N-M231 variation in Northern Eurasia.

From a climatological perspective, the evolution of N1a-F1206 and N1b-F2930 occurred close to the end of the last Ice Age and the beginning of the Holocene. As noted previously, N1b-F2930 remained in East Asia and N1a-F1206 eventually expanded northwards. This diversification of N-M231 into North Eurasian and East Asian variants conforms to an expansion model presented by Xue et al. in their 2006 study. Their report analyzes haplogroup and short tandem repeat data from the Y-chromosome. The researchers found a significant population expansion around the time of the Last Glacial Maximum in northern China. However, populations in central China started to expand much later, around the beginning of the Holocene. According to the researchers, Pleistocene populations in northern China expanded because they were able to exploit the abundant food resources on the so-called "mammoth steppes."

Much of the published data for N1a-F1206 mutations focus on the N1a-M46 "Tat" mutation and its distribution across Northern Eurasia. The most significant N1a-M46 study is that published by Ilumae et al. (2016). The researchers report that N1a-M46 evolved roughly 13 thousand years ago. This dating estimate, along with its contemporary geographic distribution, suggests that N1a-M46 is a genetic relic of the human re-expansion into Northern Eurasia around the beginning of the Holocene.

Section 3. The Reindeer Hypothesis and N1a-M46 Variation in Eurasia.

3.1. Introduction.

Ilumae et al. (2016) suggest that the rapid bi-directional expansion of N1a-M46 mutations across Eurasia was driven by the Seima-Turbino cultural phenomenon and metallurgy. However, their model is problematic because its fails to link the expansion of N1a-M46 with reproductive success, something that is clearly suggested by the Y-chromosome data. Rather, the *reindeer hypothesis*, an alternative model that incorporates anthropological and paleo-climatological perspectives, suggests that the rapid bi-directional expansion of N1a-M46 mutations across Eurasia was driven the domestication of reindeer.

3.2. Domestication of Reindeer and Reproductive Success.

Reindeer, both wild and domesticated, have been an important food resource among the cultures of Northern Eurasia for over 40 thousand years. To grasp the significance of this resource among prehistoric hunter-gatherers of the region, the reader is asked to consider the following opinion rendered by an anthropologist who specializes in this research direction:

> It [reindeer] dominated numerically and geographically, and was used by people more intensively than any other animal. It was more important than North American or Ice-Age European wild cattle, bison, mammoth, mastodon or horse. It was more important than seals and whales in all the oceans; more important than red deer, black and white-tailed deer, moose and elk. It was more important than the great African herds of antelope, zebra and gazelle. It and its hunters occupied half the land north of the equator (Gordon (2003: 15).

The behavior of wild reindeer involves migration across the tundra especially during the spring and summer. Several factors influence this behavior including the presence or absence of forage, the depth of snow, mosquitos, and the freezing and thawing of rivers (Baskin 1986). Prehistoric hunter-gatherers often intercepted herds of migrating reindeer at strategic locations, such as river crossing (e.g., Baskin 2003). While reindeer were an excellent source of protein for these hunter-gatherers, the unpredictable migration cycle of this animal meant that they were not always a dependable source of food. The domestication of this animal obviously changed this situation, and reindeer meat became a reliable year-round source of nutrition.

Gordon in his 2003 paper addresses the correlation between fertility among hunter-gatherers and the migration cycle of reindeer. According to Gordon, women require minimum of 12 percent body fat to conceive, and 18 percent body fat to carry a child until birth. Based on data he collected among hunter-gatherers that harvest reindeer, women within these groups generally conceive in the late spring and their children are born at the end of the following winter. This pattern of conception and birth among female hunter-gatherers correlates with the migration of reindeer from the forests in the early spring. They begin a trek northward onto the tundra, where calves are born and where they feed on the lichen that grows abundantly in this area. On the open tundra hunter-gatherers can easily harvest this animal, which in turn provides the nutrition need by women for conception. When reindeer return to the forest in the fall, food becomes scarcer for human societies that hunt these animals, which hinders conception among women.

The correlation between human reproductive success and the migration cycle of reindeer is significant because the domestication of reindeer broke the previous cycle of feast and famine among the Mesolithic hunter-gatherers of Northern Eurasia. A more dependable year-round source of fat and protein emerged, which meant better nutrition for the group. This, in turn, resulted in greater fertility for women and a corresponding reduction in childhood mortality. Moreover, better reproductive success, and the need to move domesticated herds of reindeer to “greener pastures” to feed a rapidly expanding population, conveniently explains the rapid bi-directional expansion of haplogroup N-M231 mutations across Eurasia

3.3. Anthropological and Climate Perspectives and Reindeer DNA.

Identifying the where and when of reindeer domestication helps to link the contemporary distribution of N-M231 mutations with greater reproductive success. It should be noted that two different models of reindeer domestication have circulated among experts in this area (e.g., Gordon 2003). The diffusion theory suggests that reindeer were initially domesticated in single location and this practice later spread through cultural contact. The evolution theory suggests domestication arose independently in several different areas. Two reports, Mirov (1945) and Gordon (2003), favor a single location near the source of the Yenisei River in the Tuva Republic of Russia. Both reports are based on anthropological data that include prehistoric reindeer pictographs found in this area.

Support for the diffusion theory stems from genetic data gathered from reindeer. Røed et al. (2008) analyzed mitochondrial and micro-satellite markers gathered from wild and domesticated herds across the Eurasian landmass. According to their data, European and Asian reindeer have a common Paleolithic origin that diversified genetically during the Holocene. Furthermore, the data reflect independent domestication of reindeer in Scandinavia and in northern Russia. The researchers, however, could not determine whether domestication in northern Russia occurred in a single location of this vast region or, alternatively, if domestication arose independently in several different locations. According to their report, much of the genetic history of reindeer in northern Russia has been erased by a long-standing practice of augmenting domesticated herds with reindeer taken from wild herds. Despite the limitations of their work, these researchers still offer an important clue as to where reindeer were first domesticated. They determined that wild reindeer from the tundra were the source of domesticated reindeer. This is an important conclusion as it eliminates a sub-species of reindeer that evolved during the Holocene in the boreal forests (or taiga) south of the Eurasian tundra. Rather, reindeer domestication occurred further north near the Arctic Circle.

The anthropological perspectives provided by Mirov (1945) and Gordon (2003), along with the reindeer genetic data provided by Røed et al. (2008), suggest that reindeer were initially domesticated in the vicinity of the Taymyr Peninsula of northern Russia. Support for this position stems from the observation that the largest herd of wild reindeer in the world is found on the Taymyr Peninsula (e.g., Pavlov 1994; Kolpaschikov 2015). Near this location in northern Russia, the Yenisei empties into the Arctic Ocean. The wild reindeer of the Taymyr Peninsula migrate northwards onto the peninsula in the spring. When autumn approaches, they migrate southwards into the boreal forest to spend the winter along the Yenisei River (Baskin 1986). Turning now to the question of when, the paleoclimatological data suggest the domestication reindeer occurred about five thousand years ago. At this point the tundra in Northern Eurasia had finally receded to the Arctic Circle (Binney et al. 2016).

Figure 15.1. Taymry Peninsula and the Yenisei River.

Leonid Khlobystin (1931-1988) was a prominent archeologist in the former Soviet Union. Much of what we know about the prehistory of Taymyr Peninsula comes from a monograph that was posthumously published in Russian in 1998 and a later 2005 English translation his work by the Smithsonian Institution. Needless to say, his research contributes enormously to deciphering the when and where of reindeer domestication. Turning now to the 2005 translation, *Taymyr: The Archaeology of Northernmost Eurasia*, researchers find a useful discussion of the environmental transition in this region that begins with a glaciated landmass roughly 20 thousand years ago. The discussion ends with the contemporary tundra ecosystem, one that sustains the large herds of reindeer found on the peninsula. This discussion is useful because it defines when reindeer appeared in the region. It was probably about 8,000 years ago, a point in climate record that follows Holocene amelioration of the extreme weather conditions of the Pleistocene.

In his discussion of the archeological record (see 2005: 24-27), Khlobystin suggests that permanent human occupation of the Taymyr Peninsula can be securely dated to about 6,000 years ago. This conclusion stems from radiocarbon data taken from Tagenar VI site. However, recent and more reliable radiocarbon data time the arrival of the Comb Ceramic Culture in Scandinavia to around 6,000 years ago (see Section 4.3 of this chapter for more details). For the purposes of this discussion, the Scandinavian data pushes the occupation of the Taymry to a point further back in time, to at least 8,000 years ago. This conclusion follows close genetic and linguistic affinity between contemporary Finns, Sami, and Nenets.

Based on the analysis of stone tools (2005:38), Khlobystin links the human occupation of the Taymyr Peninsula with the Mesolithic Sumnagin people. According to Khlobystin, this cultural tradition arose near Lake Baikal. Prior to their arrival on Taymyr Peninsula, they had occupied the

Yenisei River Basin, which is located between Lake Baikal and the Taymyr Peninsula.

Khlobystin (2005: 186-193) also discusses the evolution of the reindeer food economy found on the contemporary Taymyr Peninsula. This subsistence strategy is based on maintaining large herds of domesticated reindeer. According to the historical record, the Nenets, a Samoyed-speaking culture, are the contemporary descendants of prehistoric peoples who ultimately perfected this technique. The ability to sustain large herds of reindeer had a transitional phase that begun during the Mesolithic. According to Khlobystin, the hunter-gatherers kept a small number of domesticated reindeer that were not eaten except during a famine. Rather, they aided the hunters when they harvested wild reindeer. The domesticated reindeer provided transport, pulling sleds to and from the wild reindeer herds. Additionally, they served as decoys, allowing hunters to close the distance between themselves and the wild reindeer. The transition to large herds of domesticated reindeer was ultimately facilitated by use of dogs to guard the reindeer. Additionally, dogs were utilized to move reindeer across vast stretches of Tundra in search of forage.

3.4. Domestication of Reindeer and the N1a-M46 Mutation.

The reader is directed to **Supplementary Table 15.12**. The N1a-M46 mutation attains a significant frequency among several populations for which reindeer herding is a current or recent practice. Particularly striking about the table is that it represents an excellent cross section of populations along the entire northern Eurasian landmass: Sami (Scandinavia); Komi (Volga Uralic Region), Selkups, Nganasans, Nenets, Khanty and Dolgans (Western Siberia); Sojots, Yakuts, Evenki, Even, Yukaghir and Dolgans (Central Siberia); Tuvans and Tofalars (Southern Siberia); and Chukchi and Koryaks (Eastern Siberia).

Ancient Y-chromosome data potentially link the N1a-M46 mutation with the first domesticators of reindeer on the Taymyr Peninsula, the Mesolithic Sumnagin culture. As noted in Section 3.3. (above), the origins of the Sumnagin, are linked to hunter-gatherers who once lived in the vicinity of Lake Baikal. As shown by **Supplementary Table 15.11**, the N1a-M46 mutation was found in two remains recovered from the Lokmotive cemetery, an archeological site located in the Angara Valley about 70 kilometers downstream from Lake Baikal. One set of remains belongs to an individual who died roughly 7.6 thousand years ago. Another set dates to roughly 5.6 thousand years ago. As such, these data suggest contemporary reindeer herders inherited the N1a-M46 mutation from the prehistoric Sumnagin people.

It should be emphasized that Mesolithic hunter-gatherers with the N1a-M46 mutation and reindeer co-migrated along the Yenisei River. Then, at the Taymyr Peninsula, they encountered a geographic cul-de-sac. Previously detailed climatological and archeological data suggest this occurred about eight thousand years ago. The termination of this migration, as suggested by these data, correlates well with the beginning of reindeer domestication as suggested by an anthropological perspective of the data. According to Khlobystin and his interpretation of the archeological record (2005: 186-193), rapid population growth occurred among prehistoric peoples of the peninsula as the result of reindeer herding. Moreover, Khlobystin suggests that the transition to large scale herding required people to move their reindeer across vast stretches of the Tundra.

The position taken in the above paragraph, that the domestication of reindeer resulted in greater human reproductive success, and that this success fueled bidirectional human expansions across Eurasia, is supported not only by the archeological record, but also by contemporary N1a-M46 data provided by Ilumae et al. (2016). According to the study, informative markers, that are downstream from N1a-M46, include: N1a-B211, N1a-Z1936, N1a-M2019, N1a-VL29, N1a-B479, N1a-F4205, and N1a-B202. Their analysis of the data suggests the following: (1) they all evolved between 4,000 and 5,000 years ago; (2) their contemporary geographic distribution represents a bidirectional east-to-west or west-to-east expansion across Northern Eurasia of genetic mutations that are phylogenetically close; (3) the expansion terminated at geographic points that are distant, approximately 5,000 miles apart; and

(4), their expansion across Eurasia was rapid.

Section 4. Significance of N-M231 for Linguists.

4.1. Uralic and the *Early Farming Dispersal Hypothesis*.

Ethnologue (2018) reports 37 Uralic languages. According to the same source, around 20 million people speak a Uralic language. Roughly half this figure belongs to Hungarian. Finnish, with 5.2 million speakers, and Estonian, with a million speakers, also represent Uralic "heavyweights." The moderate to heavy frequency of downstream N1a-M46 mutations among many Uralic-speaking populations is striking. These data provide additional support for the *early farming dispersal hypothesis*, the idea that the Neolithic transformation offers a good correlation between the initial expansion of early agriculture and the current distribution of many of the world's language families. As the reader may recall, the Neolithic transformation involved farmers who cultivated crops, and pastoralists who domesticated animals. A particularly strong example of early pastoralism and language expansion occurred in Southwest Asia and North Africa. Here the domestication of sheep and goats fueled an expansion of Afro-Asiatic languages (see Chapter 10). Similarly, in northern Eurasia the domestication of reindeer fueled an initial expansion of Uralic languages. The expansion of Afro-Asiatic and Uralic languages also produced a genetic "scar." In Southwest Asia and North Africa, the J1-M267 mutation represents a genetic relic of the expansion of Afro-Asiatic (see Chapter 10). For Uralic, the genetic relic is N1a-M46.

4.2. Linguistic Relationships within Uralic.

Figure 15.2 (below) presents a traditional view of the hierarchical linguistic relationships within the Uralic language family. Campbell and Poser (2008) suggest that this tree is the product of over 500 years of comparative linguistic research that actually blazed a trail for others to follow within the field of linguistics. Moreover, both researchers stress that the Uralic is "well-studied" and "well-documented." As such, recent efforts to refine linguistic relationships within Uralic should be carefully scrutinized. In other words, linguists should think twice before abandoning the traditional tree with the "comb" model that is advocated by Ethnologue and Glottolog. See, Figure 15.3 (below). While many of the nodes in the traditional binary tree lack support from phonological reconstructions (e.g., Salminen 2007; Aikio in press), morphological tendencies among the Uralic languages arguably provide robust support for the traditional "binary" tree (see Suikkonen 2002; Janhunen 2009).

Genetic and anthropological data support the traditional "binary" Uralic tree. This model defines Samoyed and Finno-Ugric as the two main divisions of linguistic distance within the language family. As detailed in Section 3 (above), the Nenets are the closest contemporary representatives of a prehistoric population that brought Uralic languages to Scandinavia. This is linguistically significant as the Nenets language is classified within the Samoyed branch of Uralic languages. Turning back to the anthropological perspective, the Nenets live in the vicinity of the Taymyr Peninsula where reindeer were initially domesticated (see Section 3, this paper). Additionally, the Nenets have traditionally herded reindeer. Turning now to the genetic evidence, N1a-VL29 and N1a-Z1936 are among the informative N-M231 mutations identified by Ilumae et al. in their 2016 report. Both mutations are present among the Samoyed-speaking Nenets of northern Siberia. The same mutations are found among the Finno-Ugric-speaking populations Eurasia including the Finns and Sami of Scandinavia (see **Supplementary Tables 15.5** and 15.7). Thus, a synthesis of genetic, linguistic, archaeological, anthropological perspectives suggests that a proto-Uralic-speaking population lived somewhere close to the Taymyr Peninsula. The Samoyedic branch represents linguistic diversification of proto-Uralic among those that stayed. Finno-Ugric represents diversification among those who left.

Figure 15.2. Traditional Binary Tree Model of Uralic. Source: Campbell and Poser (2008: 89).

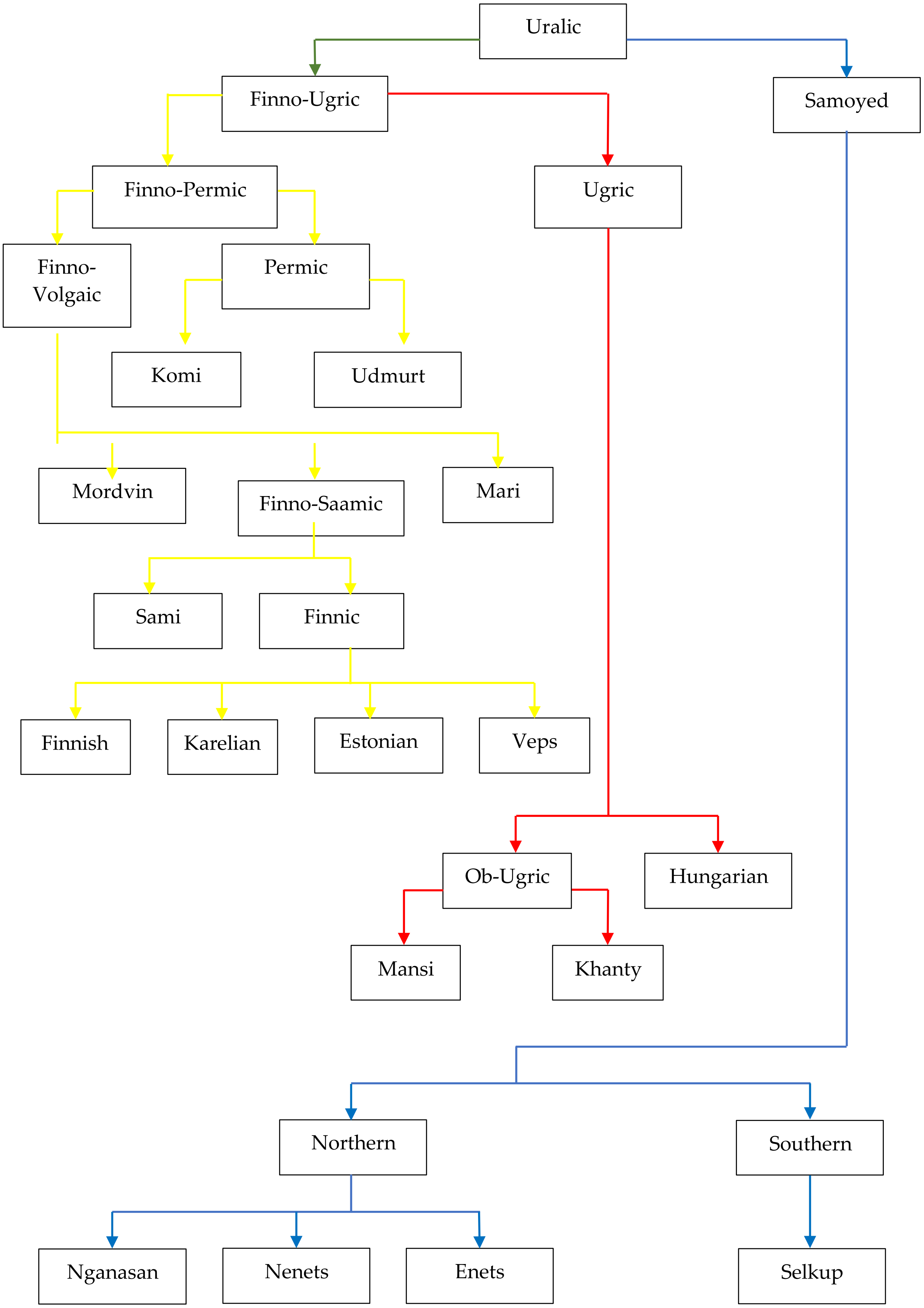
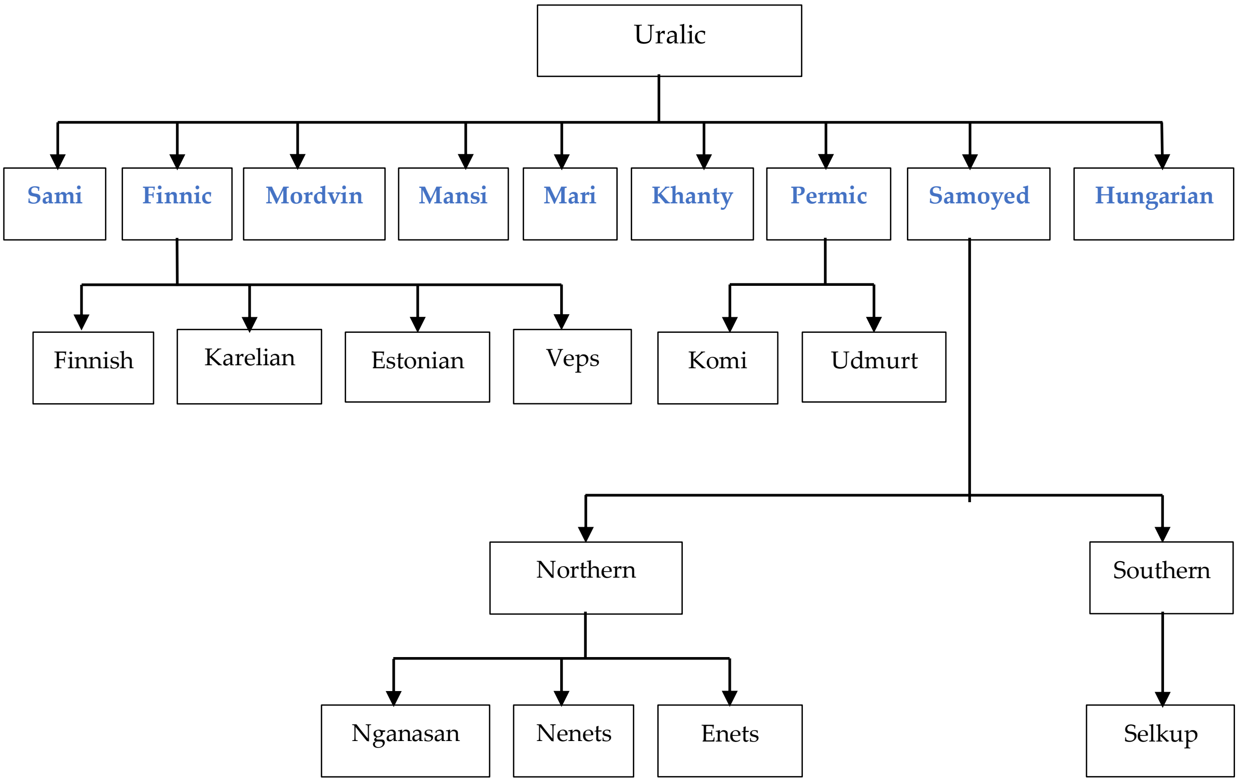


Figure 15.3. The Contemporary "Comb" Model of the Uralic Family. Source: Glottolog 4.4.



4.3. Comb Ceramic Culture.

The linguistic and genetic data clearly link the Uralic speaking populations of Scandinavia, the Finns and Sami, with those of Northern Eurasia. However, analysis of the anthropological data raises an interesting question that potentially undermines the reindeer hypothesis. Several published sources (e.g., Zvelebil 2006) suggest the Sami adopted reindeer herding only about four hundred years ago. Additionally, as noted above, genetic data from reindeer suggest independent domestication of the animal in two different regions, one in northern Eurasia, and the other in Scandinavia.

Traditionally, the arrival of Uralic languages in Scandinavia has been associated with the arrival of the Comb Ceramic Culture, which is also identified as the Pitted Ware Culture (e.g., Siiräinen 2003: 51). According to radio-carbon data, this occurred around 6,000 years ago (e.g., Piliciauskas et al. 2019). Additionally, the archeological data suggest that they practiced a subsistence strategy based on hunting and gathering, and more specifically, the harvesting of marine resources (e.g., Siiräinen 2003). Interestingly, the Comb Ceramic Culture underwent a process of "de-Neolithisation" that occurred in the Baltic prior to the expansion into Scandinavia. This conclusion follows recent analysis of settlement patterns and milk residue found on pottery as reported by Piliciauskas et al. (2020). The same study further reports that in the Baltic region, the Comb Ceramic Culture practiced cattle herding and not reindeer herding. Apparently, at some point they then decided that cattle were not worth the trouble, and they turned to the abundant marine resources of the Baltic Sea.

The reindeer hypothesis still presents a robust model of the Uralic expansion. Khlobystin (2005: 188) suggest that some of Uralic-speaking cultures adopted cattle herding during the Bronze Age. They probably encountered an ecosystem that failed to support large herds of reindeer. As discussed above, the Comb Ceramic culture experimented with cattle herding before they abandoned agriculture all together. For linguists, the above discussion of the Comb Ceramic culture helps model the prehistory of language. These data elucidate one of several trajectories that explain the contemporary pattern of global language variation. Interestingly, a reversion to foraging, as is the case with the prehistoric ancestors of Sami and Finns, is rarely observed. The Siouan-Catawban peoples and languages of the North American Great Plains might provide another example of this phenomenon (see Chapter 17).

4.4. Hungarian.

The N1a-M46 mutation maintains an astonishingly high frequency among many populations speaking a Uralic language, such as 93% of Nenets, 54% of Finns, and 31% of Estonians (e.g., Ilumae et al. 2016). However, despite speaking a Uralic language, haplogroup N-M231 and its downstream variants are almost completely absent among Hungarians (e.g., Völgyi et al. 2008). While the genetic history of Finns and Saami point to population expansion as having the potential to alter the linguistic landscape of a region, the genetic history of Hungarians reflects that population expansion is not a prerequisite for language expansion. Such a scenario for the Hungarians language agrees with the historical record. A relatively small population from Central Asia, the Magyars, invaded Europe in the fourth century. Later, a much larger Central-European population shifted to the language of their conquerors from the east.

Although Magyars contributed almost nothing to the contemporary Y-chromosome diversity among the Hungarians, a recent genetic study has identified the Mansi people of western Siberia as a potential source population for the Magyar invasion. According to Fehér et al. (2015), the N1a-L1034 mutation links contemporary Mansi and Hungarians. From the perspective of the traditional Uralic tree (see Figure 15.2 above), Mansi and Hungarian languages are two different sub-branches of Ugric, and as such, they are linguistically close. Thus, the linguistic and genetic data seem to agree.

4.5. North Germanic.

Within Scandinavia the traditional pattern of language variation consists of languages classified as North Germanic, Finnic, and Sami. N1a-M46 appears to be the genetic signature of Uralic-speaking Finns and Sami based on the moderate to heavy presence of the mutation in both populations and its comparatively low frequency among ethnic groups that speak a North Germanic language. For example, N1a-M46 is virtually absent among Danes (Sanchez et al. 2004). Among the Norwegians, less than 3 percent have the mutation (Dupuy et al. 2006). Among Swedes, the figure stands between 10 and 14 percent (Karlsson et al. 2006; Lappalainen et al. 2006). The I1a-M253 mutation, on the other hand, seems more evenly distributed among all the Scandinavian populations. It is present in about one third of Finns and Saami (Tambets et al. 2004; Lappalainen et al. 2006). Similar frequencies are detected among Danes, Norwegians, and Swedes (see Chapter 10).

The data for N1a-M46 and I1-M253 in Scandinavian populations reflects that language contact is very much a part of linguistic evolution in this area of the world. Linguistic support for this conclusion is provided by Finnish borrowings from Germanic. These data suggest cultural exchange between the Uralic and Germanic tribes of prehistoric Scandinavia. Fromm (1977), for example, argues that these loanwords point to the presence of the Germanic tribes in central Sweden during the Bronze Age, roughly 3,000 years ago. Additionally, since Finnish has changed relatively little over the two past millennia, the Germanic borrowings in Finnish are thought to provide a well-preserved image of early Germanic phonology and morphology (e.g., Loikala 1977: 229-230). Finally, the linguistic evidence also points to prehistoric contact between the Sami and the Germanic tribes (see Aikio 2020).

4.6. The Baltic.

Lithuania shares a border with Latvia, and Latvia shares a border with Estonia. Latvians and Lithuanians speak languages classified within the Baltic languages of the Indo-European language family. Estonians, on the other hand, speak a Uralic language that falls within the Finnic branch.

Laitinen et al. (2002) suggest, based on their assessment of the genetic data, that Latvians, Lithuanians, and Estonians descended from a common population based on the similar frequencies of the N1a-M46 mutation in all three populations (see, also, **Supplementary Table 15.3**). They support their conclusion by providing evidence from the archaeological record and by citing a Uralic relic found in the Latvian language. According to the study, this suggests that Estonians maintained their ancestral language, whereas Latvians and Lithuanians shifted languages, perhaps as the result of the Slavic expansion. Recent higher resolution data from Ilumae et al. (2016) confirms this hypothesis. As shown by **Supplementary Table 15.7**, N1a-VL29 represents about a third of genetic variation among Latvians, Lithuanians, and Estonians. The same data also links Baltic populations with Samoyed populations in northern Siberia, and with the Finns and Sami of Scandinavia.

4.7. East Slavic.

Russians speak an Indo-European that falls within the East Slavic branch. Among Russians, the N1a-M46 mutation potentially signals the genetic legacy of populations that shifted from Uralic to Slavic about 1,500 years ago. It should be noted that most of the N1a-M46 variation among Russians appears to be N1a-Z1936 and N1a-VL29 (see **Supplementary Tables 15.5** and **15.7**). Additionally, among ethnic Russians the frequency of haplogroup N1a-M46 is influenced greatly by geography, with a diminishing north to south frequency cline. In northern Russia about 43% of ethnic Russians have the mutation, whereas the frequency decreases to 10% in the south (Balanovsky et al. 2008). This frequency pattern is an important component of a model that explains the origins of Slavic languages as a whole. Specifically, the shift to Slavic in Eastern Europe occurred without a large population expansion from a proto-Slavic homeland. It should be noted that another key component of the Slavic expansion model is downstream variants of the R1a-M420 mutation. Additional details will follow in Chapter 18.

4.8. Altaic Languages.

For purposes of this present discussion, the term “Altaic” refers to potential areal relationships rather than a common ancestral language for Turkic, Tungusic, and Mongolic. **Supplementary Tables 15.2** and **15.3** suggest that the N1a-P43 and N1a-M46 mutations are useful markers for deciphering the linguistic prehistory of all three language families. The data from both tables also raise an interesting question, whether N1a-P43 and N1a-M46 are Paleolithic relics or, alternatively, if they define more recent population expansions associated with the *reindeer hypothesis*. N1a-M46 data for Altaic-speaking reindeer herders favor the more recent expansion (see **Supplementary Table 15.12**). Such a position is also supported by informative N1a-M46 variants for Altaic populations as a whole. At this point the reader is directed to **Supplementary Table 15.6**. The N1a-M2019 mutation appears to be an important marker for exploring the genetic history of Tungusic and Turkic-speaking population in central Siberia. The N1a-F4205 mutation, on the other hand, represents an important marker among Mongolic speakers in East Asia (see **Supplementary Table 15.8**). Finally, the N1a-B479 mutation attains a frequency of around 40% among the Nanai people, a Tungusic-speaking population of East Asia and Russia (Illumae et al. 2016).

As previously detailed in Chapter 7, the C2-M217 mutation attains a significant frequency among many of the Altaic-speaking populations. Among these populations, C2-M217 should be seen the putative Altaic marker, whereas N1a-M46 is a Uralic marker. As such, the absence of C2-M217 within an Altaic population may suggest language shift among Uralic-speaking reindeer herders. This appears to be the case among the Yakuts, Turkic-speaking reindeer herders, among whom the N1a-M46 mutation attains a frequency of 90%. Alternatively, Altaic-speaking populations with C2-M217 and N1a-M46 admixture may suggest a population history that entails assimilation of a smaller Uralic-speaking group by a larger Altaic-speaking population. This assimilation may or may not have changed the subsistence strategy within the new admixed population. Tuvans and Buryats, for example, are two Altaic populations that have significant frequencies of C2-M217 and N1a-M46. Reindeer herding has been utilized by Tuvans whereas it appears that Buryats have never utilized this subsistence strategy (Mirov 1945).

Surprisingly, a review of the available data failed to detect the presence of haplogroup C2-M217 among contemporary Uralic-speaking populations. Thus, while the data support language shift from Uralic to Turkic, a shift in the opposite direction, from Turkic to Uralic, is not supported by the available Y-chromosome data.

4.9. Altaic and Transeurasian.

Striking lexical and grammatical similarities found among the Japonic, Koreanic, Turkic, Tungusic, and Mongolic languages (e.g., Robbeets 2008) have been a topic of intense interest among linguists. The *Transeurasian hypothesis* has been formulated to explain these similarities (e.g., Robbeets 2017). An approach to this hypothesis from the perspective of historical linguistics would classify these language families as part of an Altaic or Transeurasian macro-language family (or macro-phylum). As such, linguistic similarities are explained by the evolution of Japonic, Koreanic, Turkic, Tungusic, and Mongolic from a common proto-Altaic or proto-Transeurasian language. Downrange variants of the N1a-M46 and N1a-P43 mutations attain a significant frequency among some populations that speak Turkic, Tungusic, and Mongolic languages (see **Supplementary Tables 15.2** and **15.3**). However, haplogroup N-M231 is present in only 4 percent of Koreans (Park et al. 2012), and less than one percent of Japanese (Hammer et al. 2006; Sato et al. 2014). Unfortunately, geneticists have not sequenced Koreans for informative downstream variants of the N-M231 haplogroup. Among the Japanese, the limited data from Hammer et al. (2006) suggests that most of the haplogroup N-M231 variation among this population consists of N1b-F2930, a mutation absent among Altaic-speaking populations (see **Supplementary Table 15.1**). As such, the available N-M231 data currently fails to link Japanese and Koreans with the Altaic component of the Transeurasian hypothesis.

4.10. Paleo-Siberian Languages.

The term "Paleo-Siberian" represents a convenient macro-family descriptor for several of the small North Eurasian language families including Yukaghir, Yeniseian, Eskimo-Aleut, and Chukotka-Kamchatkan as well as the Nivkh language isolate. For those interested in exploring this topic in greater detail, Vajda (2009) is highly recommended. According to the researcher, the Russian language and culture expanded into Siberia around 1582. As result, much of the indigenous linguistic and ethnic diversity of the region has been lost.

The N1a-M46 mutation (see **Supplementary Table 15.3**) attains a significant frequency among Koryaks and Chukchi, two East Siberian populations that speak a Chukotka-Kamchatkan language. Additionally, the same mutation is found among the Yupik, an Eskimo-Aleut population of the same region. Higher resolution mutations reported by Illumae et al. (2016) identify the N1a-B202 mutation as the genetic signature of these populations (see **Supplementary Table 15.9**). Finally, the N1a-M46 mutation attains a significant frequency among Central Siberian Yukaghirs. Interestingly, haplogroup N-M231 has not been detected among the Kets, the only Yeniseian population for which data is available. Similarly, haplogroup N-M231 is absent among the Nivkh (see Rootsi et al. 2007). According to Vajda (2009), both groups never adopted pastoralism but rather, remained hunter-gatherers. Perhaps this observation explains the genetic data.

The observed frequency of the N1a-B202 mutations among Yupik, Chukchi, and Koryaks provides additional support for the *reindeer hypothesis* with the idea that better reproductive success fueled a population expansion from north-central Siberia to a geographic dead-end at the Bering Sea. Further support for this idea stems from the observation that reindeer herding is practiced by the Chukchi and Koryaks. Unfortunately, Ilumae et al. (2016) were unable to identify the genetic history of the Yukaghir lower than N1a-P298 mutation. This suggests that their genetic history includes higher resolution downstream N1a-M46 mutations that have not been discovered.

Sometimes the unexpected absence of a haplogroup within a population presents useful data for researchers. Among the Yupik, who are sometimes referred to Siberian Eskimos, the frequency of haplogroup N-M231 hovers around 50% of the population (i.e., Ilumae et al. 2016). However, haplogroup N-M231 appears to be absent among North American Eskimos although they, like the Yupik, speak languages belonging to the Eskimo-Aleut language family. Founder effect and genetic drift may explain this observation. Alternatively, Paleo-Eskimos had already crossed over the Bering Sea before the expansion of haplogroup N-M231 into Eastern Siberia. Another explanation may well stem from the paucity of Y-chromosome data for Native Alaskans.

Section 5. Conclusions for Haplogroup N-M231.

The genetic, paleo-climatological, and anthropological evidence suggest that NO-M214 evolved in Central Asia around 47 thousand years ago. N-M231 then evolved from NO-M214 about 42 thousand years ago in China. Diversification within N-M231 began close to the end of the last Ice Age with the evolution of N1a-F1206 and N1b-F2930. N1b-F2930 remained in East Asia. N1a-F1206 carries the story of genetic diversity in northern Eurasia.

N1a-M46 evolved roughly 13 thousand years ago in northern China or Mongolia. Several informative downstream markers within N1a-M46 suggest a rapid bidirectional human expansion across northern Eurasia about five thousand years ago. Some researchers attribute the expansion to the development of metallurgy. However, the genetic, paleo-climatological, and anthropological evidence suggest that this expansion resulted from better reproductive success. Thus, the rapid bidirectional human expansion across northern Eurasia conforms to the *reindeer hypothesis*. Moreover, the *reindeer hypothesis* not only explains a human expansion, but also the expansion of Uralic languages. In doing so, the hypothesis provides additional support for a global pattern of linguistic evolution that follows the *early farming dispersal hypothesis*. Finally, the reindeer hypothesis elucidates the role of language contact in shaping linguistic diversity in Scandinavia, the Baltic Region, Eastern Europe, and Northern Eurasia.

It should be emphasized that hundreds of cultures and languages are dispersed across the vast Northern Eurasian landmass. The *reindeer hypothesis* hardly represents the final word for exploring the tremendous amount of cultural and linguistic diversity found in this region. Rather, the hypothesis merely represents a starting point for future linguistic research that integrates archaeological, historical, genetic, climatological, and linguistic perspectives. Future efforts to refine this language model must also consider the body of research that remains published Russian and Finnish.

Better resolution of language variation in Eurasia will require better resolution of the N1a-P43 mutation. The internal phylogeny of N1a-P43 still remains “unexplored territory.” N1a-M46, on the other hand, represents, by far, the most informative branch within N-M231 thanks to Ilumae et al. (2016) and the informative downstream mutations that they reported. Additional high-resolution sequencing of previously collected N1a-M46 samples would also be fruitful. For example, high resolution N1a-M46 data are only available for 39 Finns.

Chapter 16: Haplogroup O-M175.

Section 1. Contemporary Distribution of Haplogroup O-M175.

The reader is now directed to **Supplementary Table 16.1**. As shown by the table, haplogroup O-M175 is an especially informative marker for deciphering the population history of East Asia. Wang and Li (2013b) estimate, for example, that the haplogroup attains an astonishing frequency of about 75 percent among contemporary populations in China. Besides East Asia, haplogroup O-M175 represents an informative marker among the populations of South Asia, Island Southeast Asia, and Oceania. Additionally, as shown by the table, the O-M175 haplogroup attains a significant frequency among populations that speak languages belonging to the following language groups: Austro-Asiatic, Austronesian, Chinese, Dravidian, Hmong-Mien, Indo-Aryan, Japonic, Koreanic, Mongolian, Tai-Kadai, Tibeto-Burman, and Tungusic.

Section 2. Evolutionary History of Haplogroup O-M175.

2.1. Introduction.

The internal phylogeny of the O-M175 haplogroup is complex. A total of five supplementary figures are required to present several linguistically informative mutations associated with the haplogroup. A summation of the data suggests that O1a-M119 mutation is an informative marker for deciphering the prehistory of Austronesian languages. O1b-M95 has emerged as an especially strong marker for Austro-Asiatic languages. An especially informative marker for Japonic and Koreanic is the O1b-SRY465 mutation. The O2a-002611 mutation represents the genetic signature of Chinese languages. O2a-B451 is the genetic signature of Austronesian languages. The internal phylogeny of O2a-M117 awaits better resolution. Nevertheless, its downstream variants elucidate the prehistory of all the East Asian language families. Turning now to O2a-F114, this mutation is a lineage associated with Chinese languages. Finally, the O2a-M7 mutation is an informative marker for Hmong Mien and Austro-Asiatic languages.

2.2. Marine Isotope Stage 3.

The contemporary distribution of several haplogroups is explained by migrations across the Eurasian landmass during Marine Isotope Stage 3, roughly 50 thousand years ago. Among these haplogroups is O-M175. At this point the reader is invited to review **Supplementary Figure 1.1** from the first chapter. The NO-M214 mutation is a downstream variant of KR-M526. Dating estimates previously presented in Chapter 15 suggest that NO-M214 arose roughly 47 thousand years ago in Central Eurasia. Haplogroups O-M175 and N-M231 then evolved from NO-M214 roughly 42 thousand years ago.

As shown by **Supplementary Figure 16.1** and **Supplementary Figure 16.2**, the O-M175 haplogroup has two main branches within its internal phylogeny: O1-F265 and O2-M122. Data from Poznik et al. (2016) suggest that this initial diversification of O-M175 variation occurred roughly 32 thousand years ago just before the Last Glacial Maximum. Based on the archeological evidence associated with the success of the so-called mammoth steppe cultures, this split probably occurred in Mongolia or northern China.

2.3. Marine Isotope Stage 2.

The beginning of Marine Isotope Stage 2, roughly 26 thousand years ago, coincides with the Last Glacial Maximum. Across Eurasia deteriorating climatic conditions placed pressure on human populations to migrate southwards into refugia. For example, this change in climatic conditions may explain the expansion of C2-M217 mutations into the Russian Far East (see Chapter 7). In Western Europe, hunter-gatherers with I-M170 mutations congregated in Iberia (see Chapter 10). Similarly, the Last Glacial Maximum provides a partial explanation for the contemporary distribution of the O-M174 haplogroup and its phylogenetic "sister" clade mutation, haplogroup N-M231.

As discussed previously in Chapter 15, the contemporary distribution of N1a-F1206 in northern Eurasia suggest that even during the last glacial maximum, hunter-gatherers in northern China continued to exploit the large game resources of the "mammoth steppes." Nevertheless, archaeological, climatological, and genetic data suggest that other hunter-gatherers migrated southwards from Siberia into central China and more specifically, into refugia along the Yangtze River. Strong archaeological support for this southward expansion is the sudden appearance of micro-blade artifacts in China that occurred roughly 23 thousand years ago. According to Yi et al. (2015) the micro-blade tool making tradition was initially perfected by the mammoth steppes of Siberia, a technological innovation that met the needs of cold-adapted hunter-gatherers in this region of the world.

Turning now to the climatological perspective, a recent discussion of the Holocene transition in East Asia (Dennell et al. 2020) also identifies the Yangtze River Valley of China as a potential Ice Age refugia. A useful computer simulation study of Ice Age refugia across the Eurasian landmass (Gavashelishvili and Tarkhnishvili 2016) suggests that the terrain of this region consisted of savannah and woodlands at the time of the Last Glacial Maximum. Those who wandered into this region around the time of the Last Glacial Maximum were forced to adopt a new subsistence strategy. For these people, the large herbivore food resources of the Eurasian steppes had disappeared. Rather, small game and plant resources were now on the menu. Support for this position comes from the climate record. Dennell et al. in their 2020 paper define the Qinling Mountains of China as a geographic division between southern and northern China. This division is defined by the researchers as two different "biogeographic realms." The cool and dry steppes of northern China, the so-called "paleartic realm," supported the large herbivores. The Savana and woodlands of southern China, the so-called "Oriental realm," was considerably more humid and warmer because of the summer monsoonal rains.

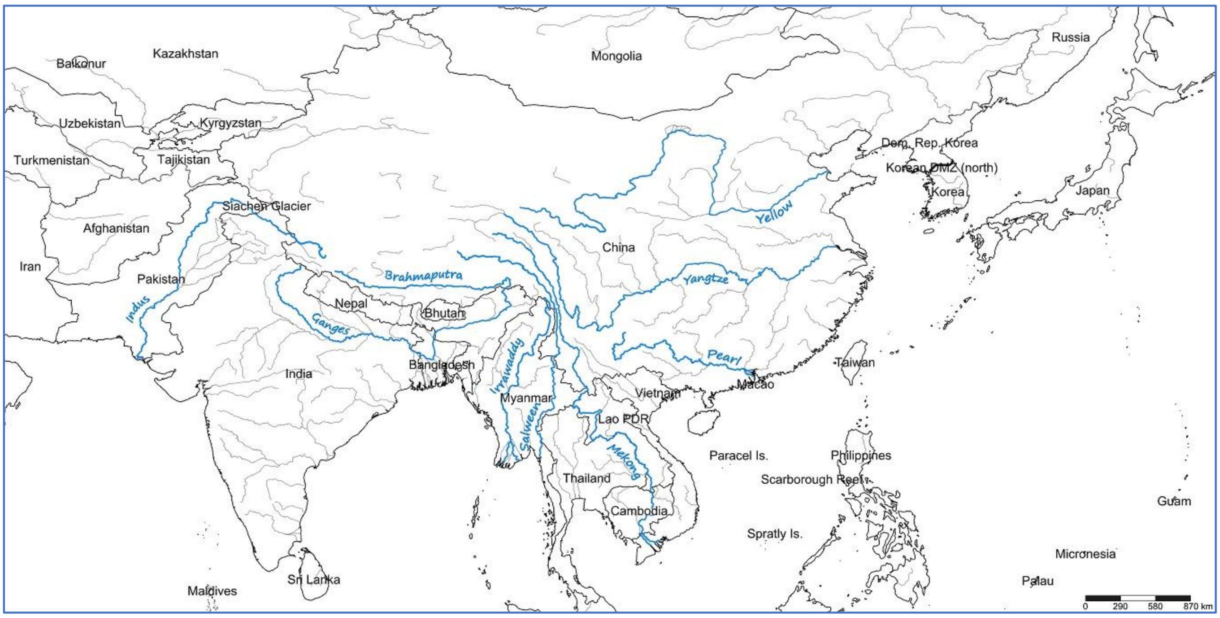
The above discussion of a climate mediated change in subsistence strategy underscores the position that people in the Yangtze River refugia were able to thrive and survive. This conveniently explains the evolution of several haplogroup O-M175 lineages near the beginning of the Holocene. O1a-M119 evolved roughly 15 thousand years ago (Sun et al 2021); O1b-M95 roughly 12 thousand years ago (Singh et al. 2021); O2a-M7 roughly 15 thousand years ago (Xia et al. 2019); and O2a-002611 roughly 13 thousand years ago (Wang et al. 2013a). The Holocene transition in southern China obviously offered opportunities and constraints for those who lived here. The evolution cultivation of rice that occurred later during Neolithic further improved reproductive success. This represents a cultural adaptation that is optimal for the monsoonal conditions of this region.

2.4. Origins of East Asian Rice Cultivation.

The Neolithic expansion of rice agriculture explains much of the linguistic diversity of East Asia, and to a more limited extent, that of South Asia and Island Southeast Asia. A particularly striking observation of contemporary demography in these regions is high population density. This observation, of course, raises an important question. How can countries such as China, India, Korea, Japan, Bangladesh, Malaysia, the Philippines, and Indonesia support such large populations? Rice agriculture provides, at least, a partial answer. As suggested previously, rice cultivation represents an optimal subsistence strategy for those living in the Oriental biogeographic realm.

The story of rice cultivation in East Asia begins about 10 thousand years ago along the Yangtze River basin of eastern China (see Figure 15.1 below). A computer simulation study from 2015 (Silva et al.) further proposes a dual origin model. This study utilized a large archaeological rice database and found that that rice cultivation may have evolved independently in the middle and as well as in lower Yangtze River Valley. According to Stevens and Fuller (2017), by around 4,500 years ago rice cultivation began to spread out of the Yangtze River Basin as the result of population pressure. Over a period of roughly 2,000 years, rice cultivation spread into southern China, and then into Southeast Asia, which includes Thailand, Vietnam, and Malaysia. Additionally, rice cultivation spread eastwards into Korea and Japan. Finally, Chinese rice cultivation spread westwards to India.

Figure 16.1. Rivers of East and South Asia.



For linguists, the development of paddy field rice cultivation is significant. Languages thrive and survive when people thrive and survive. The high population density associated with rice cultivation explains why several East Asian language families now occupy a section of the tapestry of global language variation: Chinese, Austro-Asiatic, Austronesian, Korean, Japonic, Hmong-Mien, and Tai-Kadai.

Section 3. Haplogroup O-M175 and the Origins of Chinese.

The term “Chinese” requires additional clarification at this point. In term of ethnicity, the term “Chinese” refers to the Han ethnic group. The Han are found predominately in China. They are also one of 56 different ethnic groups recognized by the Chinese government. Today they comprise almost 92 percent of the population in China, and as such, they are clearly the largest ethnic group in this country (e.g., CIA World Factbook). From a linguistic perspective, “Chinese” represents one of two main branches within the Sino-Tibetan language family (*Ethnologue* 2018). The other branch is Tibeto-Burman. Within the Chinese branch, *Ethnologue* lists 14 different languages. Furthermore, *Ethnologue* reports that around 1.3 billion people speak Chinese.

LaPolla (2001) provides an authoritative overview of the origins and expansion of Chinese languages from a linguistic perspective. First, he correlates the origins of Chinese with the evolution of rice cultivation in the Yangtze River basin. Then, he correlates the initial expansion of Chinese with the expansion of rice agriculture out of this region. As noted previously in Section 2.4, this occurred about 4,500 years ago. Finally, LaPolla explains that the contemporary distribution of Chinese also reflects internal migrations and population displacements in China as well as the rise and fall of empires and kingdoms across East Asia.

Yan et al. (2014) takes that position that 40 percent of Chinese can trace their genetic ancestry to three "Neolithic super-grandfathers." According to the study, the genetic relics of these "super-grandfathers" is the O2a-002611, O2a-M117 and O2a-F114 mutations. At this point the reader is invited to review **Supplementary Tables 16.2**, **16.3**, and **16.4**. Based on the available data, O2a-002611 and O2a-F114 represent reliable markers for deciphering the prehistory of the Chinese language branch. The O2a-M117 mutation, on the other hand, seems more problematic.

Turning now to the O2a-002611 "grandfather" mutation, Wang et al. (2013a) explored the evolutionary history of this mutation by utilizing a large data set of almost 8,000 samples. As such, their conclusions are especially persuasive. As previously mentioned, the study reports that O2a-002611 evolved about 12 thousand years ago in southeastern China. The researchers were then able to link O2a-002611 with the Han expansion during the Neolithic by provide dating estimates for the O2a-F11 mutation, a downstream variant. The estimate (about 5,000 years), along with the contemporary distribution of O2a-F11, correlate well with the co-expansion of this mutation along with Chinese and rice agriculture beginning about 4,500 thousand years ago. The expansion appears to have begun in the lower Yangtze Valley and from this region it spread northwards to northeastern China, and southwards to Laos. This study also reports data for another O2a-002611 mutation, O2a-F238, which provides a genetic relic of the Han expansion during the East Asian Neolithic.

Much of the data for the O2a-F114 "grandfather" mutation comes from Ning et al. (2016). Frequency data provided by the study suggest that this mutation is an informative marker among the Han Chinese. According to the researchers, O2a-F114 evolved roughly 8,000 years ago. However, dating estimates for several O2a-F114 downstream variants clear place its expansion during the East Asian Neolithic.

The O2a-M117 "grandfather" mutation, on the other hand, represents not only an informative marker for Chinese, but also the other East Asian languages that are discussed in this chapter (see, also, **Supplemental Table 16.3**). As such, the usefulness of this marker is limited because informative downstream markers await identification. The most reliable dating estimate for the mutation comes Wang et al. (2014). According to the study, this marker evolved roughly 7,000 years ago.

Interestingly, ancient DNA data (see **Supplementary Table 16.5**) place the evolution of O2a-M117 along the Yellow River based on the distribution of this marker and its O2a-M133 downstream variant among several different archaeological sites. A similar position can be taken with respect the O2a-F114 mutation. The O2a-002611 mutation not been detected along the Yellow River but rather at two archaeological sites that geographically distant, the Liao River Valley of northeastern China, and Man Bac in Vietnam.

Section 4. Haplogroup O-M175 and Tibeto-Burman Languages.

4.1. Introduction.

Tibeto-Burman languages are a branch within the Sino-Tibetan language family. According to Ethnologue (2018), the Tibeto-Burman branch consists of 442 languages that are organized within 12 different sub-branches. Tibeto-Burman languages are predominately found in the East Asian countries of China and Myanmar (Burma), and the South Asian countries of India, Nepal, Bhutan, and Bangladesh. A reliable estimate for the number of Tibeto-Burman speakers could not be found. The number is probably less than 100 million.

4.2. Tibeto-Burman, Linguistics, Genetics and Anthropology.

From an archaeological perspective, the starting point for a discussion of this language group begins with the Tibetan Plateau in China. While rice cultivation explains the expansion of Chinese languages, the initial expansion of Tibeto-Burman correlates well with the cultivation of barley on the Tibetan plateau beginning about 3,600 years ago. Unlike other grain crops, barley tolerates the cold and dry climate that is associated with the high altitude of this region (Zhang et al. 2016). Interestingly, barley is a crop that first cultivated in Southwest Asia about 10 thousand years. This grain may have arrived on the Tibetan Plateau via Central Asia.

An additional factor that explains the success of barley agriculture on the Tibetan Plateau is the yak, a bovine that became an essential component of the subsistence strategy in this region of the world. They provide dairy products, meat, hide and transport. A genetic study from 2015 (Qiu et al.) compared data from wild and domesticated yaks. The researchers report Yaks were domesticated about 7,300 thousand years ago, roughly 3,000 years before the arrival of barley agriculture in the region.

In addition to the cultivation of barley and the domestication of the yak, an evolutionary adaptation also explains the success of Tibeto-Burman languages. The Tibetan Plateau lies at an average altitude of 4,000 meters above sea level. Here, hypoxia and altitude sickness pose a significant health danger. People from lower altitudes can, over time, become acclimated to living at high altitude. Nevertheless, Tibetans have an evolutionary adaptation that allows them to utilize the depleted oxygen level more efficiently than those who have moved to Tibetan from a lower altitude (Wu and Kayser 2006). A study from 2017 (Yang et al. 2017) focused on nine different sections of the human genome (or loci) that potentially control this evolutionary adaptation. The study compared the genomes of about 3,000 Tibetans with 7,000 non-Tibetans from East Asia. This comparison indicates that the Tibetans and Han separated about 4,700 years ago, which is consistent with the Y-chromosome and archaeological data.

As discussed previously in Chapter 4, a particularly significant genetic characteristic of Tibetans and the Tibetan Plateau is the elevated frequency of haplogroup D-M174. According to Qi et al. (2013) around 54 percent of Tibetans have the mutation. Moreover, the mutation represents a genetic relic of hunter-gatherers that wandered into this region during the Paleolithic. The other significant haplogroup among Tibetans is, of course, O-M175, which is present in about 33 percent of Tibetans males (Qi et al. 2013).

The expansion of O-M175 onto the Tibetan Plateau potentially signals the arrival of the Yangshao culture in the region about 7,000 years ago. This follows linguistic and anthological interpretations of the data that posit a population split of proto-Chinese and Proto-Tibeto-Burman speakers on central plains of Yellow River and their migration to the Tibetan Plateau along this waterway (LaPolla 2013; Zhang et al. 2016). A similar position is also taken by the geneticists and their interpretation of the contemporary O2a-M117 data (Kang et al. 2012; Wang et al. 2014). It should be noted that the O2a-M117 mutation represents around 90 percent of the O-M175 variation among the Tibetans (Qi et al. 2013). Additionally, ancient O2a-M117 mutations have been recovered from archaeological sites along the Yellow River (see **Supplementary Table 16.5**).

4.3. Southward Expansion of Tibeto-Burman.

As stated earlier, Tibeto-Burman languages are also found in South Asia. It appears that these languages expanded into the region as the result of population pressure on the Tibetan Plateau. Unlike the Tibetan Plateau however, haplogroup D-M174 attains a low frequency among the Tibeto-Burman speaking populations of India, Nepal, Bhutan, and Bangladesh (e.g., Sahoo et al. 2006; Sengupta et al. 2006; Trivedi et al. 2008, Gazi et al. 2013, Tamang et al. 2018). Rather, haplogroup O-M175 and the downstream O2a-M117 mutation point to Tibet as the source of Tibeto-Burman languages that are found in South Asia. Unfortunately, much of the South Asian data reports frequency results for poor resolution markers that are upstream from O2a-M117. One exception is Debnath et al (2011). They report O2a-M117 frequencies between 25 and 42 percent for Tibeto-Burman-speaking populations in Eastern India. Gayden et al. (2007) is another exception. They report O2a-M117 frequencies between 21 and 84 percent for Tibeto-Burman-speaking populations in Nepal.

It should be noted that the Tibeto-Burman languages of Myanmar are classified within the Ngwi-Burmese sub-branch. Historical evidence suggests the expansion of Ngwi-Burmese from the Tibetan plateau is unrelated to the expansion of Tibeto-Burman sub-branches found in South Asia, such as Central Tibeto-Burman, Sal, and Western Tibeto-Burman. Rather than an agricultural expansion, Ngwi-Burmese correlates better with the rise and fall of the Pyu civilization and their migration along the Salween River (see La Polla 2013: 206-207).

4.4. Trans-Himalayan.

Linguists should be aware of a proposed "Trans-Himalayan" language classification which appeared in a 2018 genetic study of South Asia (Tamang et al.). The traditional linguistic view of Tibeto-Burman and Chinese languages (e.g., *Ethnologue*) assigns these language groups as the two main branches of the Sino-Tibetan language family. However, the linguist George van Driem contests this arrangement. He views Chinese (or Sinitic) as a branch or sub-group of Tibeto-Burman (Driem 2005). He also proposes a Trans-Himalayan phylum that contains many of the Tibeto-Burman languages (see Driem 2014).

The Trans-Himalayan model is controversial (e.g., LaPolla 2016) and it is not the goal of this discussion to endorse the hypothesis. Rather, the purpose is to illustrate a potential flaw with the Sino-Tibetan family classification. It seems to be an unnatural division of the data, especially from an anthropological perspective. The origins and expansion of Chinese languages stem from the cultivation of rice and the associated phenomenon of high population density. The origins and expansion of Tibeto-Burman languages, on the other hand, partly stems from the success of barley cultivation on the Tibetan Plateau. As such, a more natural division of the data might entail the creation of a Tibeto-Burman language family and a separate Chinese language family.

Section 5. O-M175 and Austronesian.

5.1. Introduction.

The Austronesian language family occupies a large corner of the global linguistic tapestry with over 1,200 languages and 324 million speakers (*Ethnologue* 2018). The Austronesian language family has two main branches, Formosan and Malayo-Polynesian. Formosan consists of 20 languages found on the island of Taiwan. Malayo-Polynesian, on the other hand, consists of 1,236 languages that have a north to south geographic distribution from Hawaii to New Zealand, and a west to east distribution from Madagascar to Rapa Nui (Easter Island).

5.2. The Lapita Cultural Expansion.

The Formosan branch of the Austronesian language family represents the linguistic signature of Taiwanese aboriginals as well as a linguistic relic of the prehistoric Dapenkeng culture. The Dapenkeng migrated to Taiwan from the East Asian mainland about 5,500 years ago. For about 700 years the Dapenkeng were hunter-gatherers. Their subsistence strategy included the harvesting of marine resources. Then about 4,800 years ago they adopted agriculture and began to cultivate foxtail millet and rice (Hung and Carson 2014). About 4,000 years ago, as the result of soil depletion and population pressure (Bellwood 2005: 135), rice agriculture spread from Taiwan to the Philippines.

Linguistically, this expansion triggered a split in the Austronesian language family and the evolution of Malayo-Polynesian branch. From an archaeological perspective, the expansion follows the migration of the Lapita culture, a term derived from a unique style of pottery. The Lapita culture initially expanded southwards through the Philippines to Borneo. From Borneo, around 3,400 years ago (Bellwood 2005: 137), a second Austronesian expansion occurred, with some migrating westwards in the direction of Malaysia, while others migrated eastwards in the direction of New Guinea.

From a genetics perspective, two different branches within the O-M175 phylogeny, are especially helpful for deciphering the initial expansion of Austronesian, O1a-M119 and O2a-N6. Within O1a-M119, the O1a-M307 and O1a-M110 mutations are the most informative. Within O2a-N6, the most informative mutation is O2a-B451. For additional information, the reader should review **Supplementary Figures 16.1** and **16.2** as well as **Supplementary Tables 16.6**, **16.7**, and **16.17**.

5.3. Eastward Expansion of Austronesian into Oceania.

According to Bellwood (2005: 134-141) around 3,000 years ago the Lapita culture began to spread across eastern Indonesia and Papua New Guinea (see, also, Horsburgh and McCoy 2017). By around 2,000 years ago the Lapita culture reached western Oceania. Finally, by around 1250 AD, after colonizing many of the islands of central and eastern Oceania, the Lapita cultural expansion terminated in New Zealand. The genetic picture of this secondary expansion is rather interesting and complicated. To draw a simpler picture, one could compare the Austronesian expansion to a city bus with a long route that begins in Taiwan and ends in New Zealand. In Taiwan, passengers with the O1a-M307, O1a-M110, and O2a-B451 mutations started the journey. On New Guinea, passengers with the C1b-M208 mutation climbed aboard the bus. In western Oceania, the passengers from Taiwan reached end their journey. The passengers from Papua New Guinea rode the bus to the end of the line. The more complicated picture of the eastward Austronesian expansion points to the survival of cultural continuity despite population replacement, and the leveling of genetic diversity as the result of founder effect.

The Austronesian colonization of eastern Indonesia and Papua New Guinea is reflected by the frequency pattern of the O1a-M119 and O2a-B451 mutations in Island Southeast Asia. After the Austronesians arrived in this region, admixture occurred between this group and the Papuan-speaking populations. This is reflected by C1-M208, M-P256, and S-B254 mutations found in the Austronesian-speaking populations of the region and the fact that these mutations are the genetic signature of Papuan-speaking populations who had inhabited Island Southeast Asia for about 50 thousand years prior to the arrival of Austronesians. Finally, the genetic data suggest that a new population carried Austronesian eastwards into Oceania

According to Horsburgh and McCoy (2017) only a third of the Y-chromosome variation in Polynesia has a potential East Asian/Taiwanese origin. The remaining variation originated on New Guinea. Interestingly, based their analysis of mitochondrial DNA, which provides a maternal genetic perspective, the genetic history of Polynesian women is almost entirely linked to East Asia and Taiwan. This asymmetrical picture of population origins requires additional attention in the future. Perhaps this asymmetry simply reflects the effect of genetic drift and founder effect in Oceania. Such a conclusion is consistent with the data from New Guinea which reflect East Asian and Papuan female admixture. Kayser et al (2008), for example, report data from the Admiralty Islands of Papua New Guinea. Around 40 percent of the mtDNA has Papuan origins and the remainder has an East Asian origin. Another example comes from Delfin et al. (2012) and data for the Solomon Islands. The Papuan contribution is potentially 22 percent, and the remainder is East Asian.

5.4. Expansion of Austronesian into Western Indonesia and Malaysia.

Previously in Chapter 13: Section 1, the concept of the so-called Wallace Line was discussed.

The division was initially introduced to describe botanical features that are unique to Island Southeast Asia versus those unique to East Asia (see Blust 2013: 6-7 for additional information). See, also, Figure 6.1 from Chapter 6). Today this term conveniently delineates western Indonesia from eastern Indonesia. Additionally, the Wallace line is important for anthropology as one finds human phenotype differences on both sides of the divide. Finally, a study from 2010 (Karafet et al.) found significant genetic differences between those living west of the division compared to those living on the eastern side. Haplogroups M-P256, S-B254, C-M38 and K-M526* represent an especially strong Papuan component of populations in eastern Indonesia. However, this component is weak or absent in western Indonesia. Rather, they have a strong East Asian component, the O1b-M95 and O2a-M7 mutations (see **Supplementary Tables 16.9** and **16.10)**. In eastern Indonesian, this East Asian component is essentially absent. Turning now to potential genetic signatures of an Austronesian component in Indonesia, it should be noted that data from Karafet et al. (2010) suggests that the Austronesian contribution hovers around 50 percent in western Indonesia, whereas the figure is around 13 percent for eastern Indonesia. This conclusion was extrapolated from the frequency results for the O1a-M119, O1a-M307, O1a-M110, and O2a-P201 mutations.

The above discussion of the Wallace line helps to provide geographic and genetic context to a westward expansion of Austronesian languages from Borneo roughly 3,400 years ago. Malaysians and Indonesians predominately speak Austronesian languages. However, Austro-Asiatic languages remain part of the linguistic tapestry found in this area of the world. Moreover, Austro-Asiatic languages arrived in the region just before the Austronesians. The archeological record supports such a scenario. According to Bellwood (2005:139), rice agriculturalists from Thailand migrated onto the Malay Peninsula roughly 4,500 years ago. Such a scenario is also consistent with genetic data. According to Arunkumar et al. (2015), the O1b-M95 mutation reflects a southward expansion of Austro-Asiatic-speaking rice farmers from Laos around this time. Data from Karafet et al. (2010) and Arunkumar et al. (2015) further suggest that after the Austronesians arrived, admixture occurred between this group and Austro-Asiatic populations. Thus, part of the story of Austronesian languages in western Indonesia and Malaysia appears to entail language shift from Austro-Asiatic to Austronesian. Recent analysis of Malay immigrants in Singapore (Sun et al. 2020) also supports this model.

5.5. Westward Expansion of Austronesian into East Africa.

Linguistic diversity in East Africa is generally associated with languages that fall within the Afro-Asiatic, Nilo-Saharan, or Niger-Congo language families. Thus, it is rather unexpected to find Austronesian languages on the island of Madagascar. Linguistic, historical, and genetic data provide an explanation.

The Great Barito sub-branch of Austronesian places the origins of Malagasy languages somewhere in the vicinity of Indonesia. Ethnologue (2018) lists 35 Great Barito languages. Twenty-three of these languages are found either in Malaysia, Indonesia, or the Philippines. The remaining languages are classified within a Malagasy branch of the Great Barito and are spoken thousands of kilometers away off the eastern coast of Africa. Of these 12 languages, one is spoken on the Island of Mayotte and the remaining are spoken on the island of Madagascar.

According to the archaeological and historical record (e.g., Blench 2010), Madagascar was initially settled by hunter-gatherers from East Africa about 4,000 years ago. Austronesian contacts with East Africa began around 2,000 years with the arrival of ships from Malaysia. The Malays periodically conducted raids. They also traded extensively with the East Africans, perhaps to obtain cinnamon. The East Africans, in turn, might have received chickens, bananas and taro root. Finally, around 1,500 years ago, after centuries of raids and trade, the Malays established permanent settlements on Madagascar.

To discuss the genetic data, it is important to emphasize that the term “Malagasy” also has an ethnic connotation and describes the inhabitants of Madagascar. According to the genetic data, the Malagasy people are a blend of populations from East Africa, the Middle East, and Island Southeast Asia (Capredon et al. 2013; Poetsch et al. 2013; Tofanelli et al. 2009a). Based on the frequency of the O1a-M110 and O1b-M111 mutations, the Austronesian contribution among contemporary Malagasy is about 20 percent.

The Ma'anyan people represent a potential source population for the Malagasy language based on the linguistic evidence. An interesting study from 2015 (Kusuma et al.) explored a potential genetic connection between the Malagasy and the Ma'anyan people of Borneo. The study failed to find any evidence of a close genetic relationship between the two groups. Rather, the genetic evidence can only pinpoint Malagasy origins within a region, either Indonesia or Malaysia. Their findings seem consistent with the anthropological data provided by Blench (2010). He reports the Great Barito languages came from Indonesians who were pressed into service onboard the Malay ships. As such, it appears that the Austronesian settlers of Madagascar represented several different Indonesian and Malaysian ethnic groups that utilized a common Great Barito language as a *lingua franca*.

5.6. The Austronesian Advantage.

Based on the number of speakers and its vast geographic distribution, Austronesian is indeed a significant linguistic "heavyweight." Donohue and Denham (2010) provide a useful summation of approaches and issues with respect to the history of Austronesian languages. They take a position that correlates the success of Austronesian with trade network that were controlled by Austronesian-speaking populations. These trade networks flourished because technological advantages, such as outrigger canoes. The also flourished because of good navigational skills that eventually carried the Austronesians over vast stretches of open water (for a more detailed discussion, see Blust 2013: 11-17).

Correlating the so-called success of Austronesian with trade is problematic because trade does not necessarily produce a reproductive advantage. In other words, Austronesian behaves much like language families that co-expanded with early agriculture, such as Niger-Congo or Sino-Tibetan. These languages thrive and survive because agriculture supports much higher population density than hunting and gathering. Bellwood (2005: 141) explains the importance of agriculture and its role in the Austronesian expansion. He writes that most Austronesian-speaking populations practice agriculture and without it, the Austronesians could not have colonized Oceania. In short, agriculture appears to be a far more crucial component of the Austronesian success story than just trade networks and technology.

As explained earlier, the Austronesian agricultural expansion began with rice and millet cultivation on Taiwan. When the Austronesian expansion reached Borneo, climatic conditions no longer supported the cultivation of grain crops. At this point the Austronesians began to cultivate tubers and tree crops that flourish in Island Southeast Asia and Oceania. Tubers include taro and yams. Examples of tree crops are sugar cane, bananas, pandanus, breadfruit, sago palm, canarium nuts and coconuts (see Bellwood 2005: 130-139); and Blust 2013: 6-7 for additional details). Nevertheless, correlating the success of Austronesian with tuber and tree crop agriculture also seems problematic. Papuans also cultivate these crops (see Chapter 13). Why, then, would agriculture have been hugely successful for Austronesians, and moderately successful for Papuans?

As previously detailed in Chapter 13, the Papuans of New Guinea congregated in the central highlands of this island. Coastal farming on the island coincided with the arrival of the Austronesians roughly 3,000 years ago. As such, Malaria could explain the differing Papuan and Austronesian patterns of early agricultural activity on New Guinea (see Kelly 1990). During the Holocene, the Papuans may have moved to the highlands to avoid malaria. Austronesians, on the other hand, were able to farm the coastal areas of the island because of an evolutionary adaptation that made them resistant to tropical splenomegaly syndrome, a massive and fatal enlargement of the spleen that occurs as the result of chronic exposure to malaria. Evidence for this so-called "Austronesian advantage" stems from a study published by Clark and Kelly in 1993. The researchers compared gamma globulin polymorphisms from Austronesian and non-Austronesian populations on New Guinea. Gamma globulin was examined because the marker has a strong association with the immune system. The researchers were able to identify a specific polymorphism characteristic of lowland Austronesian-speaking populations who are resistant to tropical splenomegaly syndrome. They also identified another polymorphism associated with highland Papuan groups, populations that are highly susceptible to tropical splenomegaly syndrome.

Clark and Kelly offer several points that are useful for understanding the Austronesian advantage. Anopheles mosquitos are the "vector" that transmits *Phasmodium*, the parasite that causes Malaria (see Cox 2010 for more details). These mosquitos thrive in the wet and swampy lowlands of New Guinea, whereas they are far less prevalent in the highlands. Furthermore, lowland coastal agriculture furthers intensifies the spread of malaria by creating habitat that facilitates the breeding cycle of these mosquitos. Finally, lowland coastal agriculture creates permanent human settlements which provide a host population for the *Phasmodium* parasite. The "Austronesian advantage" suggests that Austronesians could farm the coastal areas of New Guinea, whereas such activity for Papuans would have been lethal.

Since the prevalence of malaria diminishes with altitude, malaria avoidance would conveniently explain why the Papuans occupied New Guinean highlands at the onset of the Holocene. This presupposes the presence of malaria on New Guinea before the arrival of the Austronesians. Such an argument seems plausible based on a recent study of the *Plasmodium vivax* organism, a species of the *Plasmodium* parasite that is especially prevalent in the so-called Austronesian world. Loy et al. (2018) report a close genetic relationship between *Plasmodium vivax* parasites that infect chimps and gorillas in Africa and the *Plasmodium vivax* parasites that infect people in Island Southeast Asia. *Plasmodium vivax* parasites from both regions share a common ancestor that evolved in Africa. When people migrated out of Africa around 100 thousand years ago, the *Plasmodium vivax* organism essentially "hitched-a-ride" with the humans.

Clark and Kelley (1993) also suggest that admixture between Austronesians and Papuans created a new population that inherited a genetic resistance to tropical splenomegaly syndrome. Consistent with recent Y-chromosome evidence, their data suggest that some Papuans eventually joined the lowland Austronesian farming settlements on New Guinea. Papuans and Austronesians then produced children. Their children then inherited an evolutionary adaptation to tropical splenomegaly syndrome.

Section 6. O-M175 and Austro-Asiatic.

6.1. Introduction.

Based on data from *Ethnologue* (2018), the Austro-Asiatic language family consists of 167 languages spoken by around 105 million people. These languages are distributed across South and East Asia, primarily in India, Bangladesh, Myanmar, Cambodia, Laos, Thailand, Vietnam, and Malaysia. Within the Austro-Asiatic language family, the Munda and Mon-Khmer branches form the two main divisions. Munda represents the Austro-Asiatic languages of South Asia and Mon-Khmer represents the Austro-Asiatic language of East Asia. Among the Austro-Asiatic language, Vietnamese has attained official language status in Vietnam and Khmer is the official language of Cambodia.

6.2. Anthropological and Linguistic Perspectives.

Efforts to identify the putative homeland of Austro-Asiatic languages have produced three different models that place the origins of this language family either in India, the Mekong River Valley of Laos, or southern China. To discuss the Mekong River Valley model, it should be noted that the Mekong is among the major waterways of East Asia. It flows over 4,000 kilometers from the Tibetan Plateau through Yunnan province into Myanmar, Laos, Thailand, Cambodia, and Vietnam, where it empties into the South China Sea. The Mekong River Valley is located where the borders of Myanmar, Laos and Thailand converge on map. The linguist Paul Sidwell (2010) has identified this area as the geographic point of origin for Austroasiatic languages. He favors the Mekong River Valley with the idea that the region of greatest linguistic diversity also defines the geographic origins of a language family. Sidwell builds his argument through an analysis of the morphological, phonological, and lexical data.

George van Driem in a paper he published in 2011 advocates the India model of Austro-Asiatic origins based on phonological reconstructions. According to Driem, Austro-Asiatic originated in India because linguistic reconstructions point to a hot and humid tropical climate not found in southern China. Driem also offers botanical evidence. He suggests that modern domesticated rice originates from a hybrid of three different Neolithic variants, indica, japonica, and dry upland rice. He asserts that an initial hybrid of indica and dry upland could have only occurred somewhere near the Bay of Bengal. However, this suggestion conflicts with a more authoritative analysis. Fuller (2012) suggests that modern domesticated rice is a hybrid of just two variants, proto-indica from India and domesticated japonica from China.

The southern China model identifies the Three Gorges Region of the Yangtze River as the putative homeland of Austro-Asiatic languages. Higham (2002), based on his interpretation of the archaeological and linguistic evidence, argues that Munda and Mon-Khmer split about 6,000 years ago in this region of Sichuan province. Austro-Asiatic languages and rice agriculture then expanded upstream along the Yangtze into the Yunnan province of southwestern China. Austro-Asiatic and rice agriculturalists then expanded out of Yunnan along major river systems. According to Higham (2002), river systems were utilized to avoid travel through the dense forest canopy. Munda and rice agriculture eventually expanded into northeastern India along the Brahmaputra River. Meanwhile, Mon-Khmer and rice agriculture radiated southward from Yunnan along several major river systems including the Mekong, Irrawaddy, Chao Praya, and Red.

Higham (2002) provides linguistic support for this model by offering proto-Austro-Asiatic reconstructions for terminology related to rice agriculture. He also cites Mon-Khmer languages found in China, especially those that fall within the Palaungric sub-branch. One of these languages is P'uman, which is spoken by the Blang ethnic group in Yunnan province. According to Higham (2002), P'uman has the distinction of being the northernmost Austro-Asiatic language. Furthermore, its location on the Mekong River supports a close correlation between the expansion of early rice agriculture and the expansion of Mon-Khmer languages via this waterway.

Focusing now on the southern China model and the Munda language branch, Zhang and Hung (2008) date the arrival of rice agriculture in Yunnan at around 4,000 years ago. Fuller (2012), like Higham (2002), suggests that shortly after the arrival of rice agriculture in Yunnan, japonica rice advanced westward into India along the Brahmaputra River and eventually onto the Ganges plain. It should be noted that the cultivation of proto-indica and other rice strains in India predate the arrival of japonica. However, pre-japonica rice agriculture in India was characterized by the casual dry land cultivation of a grain that was rotated with other crops. As such, pre-japonica rice agriculture never became a significant source of food in India (Fuller 2012; Bates and Singh 2017). Rather, rice only became a significant food staple about 3,000 years ago, about a thousand years after the introduction of the japonica variety from China. By this time, japonica and indica had been developed into a high-yield hybrid rice. Furthermore, by this time farmers in India had perfected rice paddy cultivation, a technique that fully exploits the potential of rice agriculture. Thus, the arrival of Munda languages and japonica rice in India substantially altered the demographic landscape with the introduction of a rice variety that sustains high population density.

6.3. Genetic Perspectives.

Turning now to the genetic evidence, the O1b-M95 mutation attains a significant frequency among the Munda and Mon-Khmer populations of Asia (see **Supplementary Table 16.9**). Based on this observation, some researchers suggest the mutation stands as a strong genetic relic of the Austro-Asiatic expansion (e.g., Chaubey et al. 2011). However, this position is undermined by the strong frequency pattern of O1b-M95 that is observed in populations that speak Hmong-Mien, Tai-Kadai, Austronesian, and Sino-Tibetan languages.

Efforts to understand the evolutionary history of the O1b-M95 mutation were hampered over uncertainty as to when evolved. For example, Kumar et al. (2007) suggest that the mutation evolved about 65 thousand years ago. A study from 2015 (Zhang et al.) suggests that mutation arose between 20 and 40 thousand years ago. The most reliable dating estimate, based on methodology and the general pattern of early Holocene haplogroup O-M175 diversification, is provide by Karmin et al. (2015) and Singh et al (2021). Both studies suggest that O1b-M95 evolved roughly 12 years ago.

Defining where O1b-M95 evolved also remains problematic. Kumar et al. (2007) suggest eastern India. Zhang et al. (2015) suggest southern China. Arunkumar et al. (2015) suggest Laos. Ancient DNA data from Li et al. (2007) supports the idea that O1b-M95 evolved along Yangtze River and the Three Gorges Region (see, also, **Supplementary Table 16.5**). Taking this a step further, these data suggest that the Yangtze River became a corridor that carried Austroasiatic languages into Yunnan province and beyond, as suggested by southern China model.

Another barrier that has hindered analysis of the O1b-M95 mutation is the identification of informative downstream mutations. The recent publication of four studies (Kutanan et al. 2019, Kutanan et al. 2020, Macholdt et al. 2020; and Singh et al. 2021) has greatly elucidated its internal phylogeny. As shown by **Supplementary Figures 16.3** and **16.4**, O1b-F1252 and O1b-M1283 form the two main O1b-M95 lineages. Moreover, Singh et al. (2020) identify a downstream variant of the O1b-M1283 mutation, the O1b-B418 mutation, as the genetic signature of Munda speaking populations in India. As such, the O1b-F1252 and O1b-M1283 bifurcation of O1b-M95 mimics the bifurcation of Austro-Asiatic into the Mon-Khmer and Munda branches.

Section 7. Tai-Kadai and Haplogroup O-M175.

7.1. Introduction.

The Tai-Kadai language family is also known as Kra-Dai and Daic. According to *Ethnologue* (2018), about 81 million people speak a Tai-Kadai language. This language family consists of 91 different languages and is found in China, Thailand, Vietnam, Laos, and Myanmar. A handful of Tai-Kadai languages are also found in India. The Tai-Kadai language family has three main internal branches: Hlai, Kam-Tai and Kra. Hlai consists of two languages found on Hainan Island in China. Kra has 16 languages that are found in China and Vietnam. The remaining 72 languages belong to the Kam-Tai branch. Among the Kam-Tai languages, Thai, with around 60 million speakers, attains official language status in Thailand.

7.2. Tai-Kadai Homeland.

Paul Sidwell, a linguist who specialized in the languages of East Asia, places the origins of the Tai-Kadai family in southern China. He further suggests a southward expansion of these languages beginning around 2,500 years ago (see Sidwell 2013). From an archaeological perspective, Zhang and Hung (2012) raise the possibility that Tai-Kadai represents a linguistic relic of hunter-gatherers who eventually adopted farming and rice cultivation. A genetic study (Brunelli et al. 2017) takes the same position. Blench (2013) notes, however, that the prehistory of Tai-Kadai remains somewhat murky. He suggests that the historical kingdom of Siam best explains the position attained by Tai-Kadai languages within the contemporary tapestry of global language variation.

Although the genetic, linguistic, and archeological evidence point to mainland China as the putative homeland of Tai-Kadai, a genetic study from 2008 (Li et al.) suggests Hainan, an island in the South China Sea and China's southernmost province. As noted earlier, the Hlai branch of Tai-Kadai consists of two languages spoken on Hainan Island. Li et al. (2008) take the position that Hlai is a linguistic relic of the earliest Tai-Kai languages that it evolved among the aboriginal populations of the Island. Their arguments stem primarily from analysis of the elevated frequency of O1a-M119 and O1b-M95 mutations among the Hainan aboriginals.

From a linguistic perspective, it seems just as plausible that Hlai evolved from a Tai-Kadai language once spoken in mainland China (Norquest 2007). Then, at some point, speakers of early Hlai made a sea crossing to the island. Interestingly, a recent study from 2021 (Sun et al.) raises the possibility that Tai-Kadai evolved in southeastern China along the coastline of the South China Sea. The study involved phylogenetic analysis of mutations that are downstream from O1a-M119. Among these downstream mutations, O1a-F4084, O1a-K644, and O1a-K644 were identified as putative markers among Tai-Kadai-speaking populations (see **Supplementary Figure 16.5** for additional information). The idea that Tai-Kadai evolved along the coast of southeastern China is also supported by ancient O1a-M119 harvested from Neolithic archaeological sites in the Yangtze River Delta (see **Supplemental Table 16.5**).

7.3. Tai-Kadai and Austronesian.

Among the linguists, some researchers have explored the possibility that Tai-Kadai and Austronesian are united by a common proto language (see Ostapirat 2018 for an overview). Another researcher takes the position that Tai-Kadai is a sub-branch within the Austronesian language family (Sagart 2004). Thurgood (1994), on the other hand, offers compelling arguments for a borrowing relationship between Austronesian and Tai-Kadai somewhere in Guizhou and Guangxi provinces of southern China about 4,000 years ago. A genetic perspective (Sun et al. 2021), based on phylogenic relationship downstream from O1a-M119, presents evidence of an early relationship between Austronesian and Tai-Kadai. Such an argument is potentially supported by a study that places the origins of Austronesian in mainland China based on analysis of the O2a-N6 mutation (Wei et al. 2017a). Nevertheless, the weight of archaeological, linguistic, and genetic data, as presented above in Section 5, offers persuasive arguments for Taiwan as the putative homeland of Austronesian languages.

Rather than an early prehistoric relationship, whether it be a common proto language or language contact, Austronesian influence on Tai-Kai may well stem from the historical rise and fall of the Austronesian-speaking Champa civilization, which Sidwell (2013) dates between 500 BC and 1500 AD. A similar position was taken by Doi (2012) and his analysis of Austronesian influences found in Austro-Asiatic languages, especially Vietnamese. An extension of Doi's argument suggests that an intense language contact relationship not only existed between the Champa and Austro-Asiatic populations, but Champa and Tai-Kadai populations.

Focusing now on the genetic evidence for Champa influence, Li et al. (2013) report Y-chromosome data for the Utsat people who are potential relic population. Their results suggest that the Austronesian expansion onto mainland East Asia was carried by a small population from Island Southeast Asia. This small Austronesian-speaking population then admixed with a larger non-Austronesian-speaking mainland population. A similar conclusion was reached by He et al. (2012) in their study of ethnic Cham and Kinh in Vietnam as well as Thais and Laotians.

Section 8. Hmong-Mien and Haplogroup O-M175.

8.1. Introduction.

The contemporary distribution of Hmong-Mien languages is found among the hills of southern China, northern Laos, and northern Vietnam. According to *Ethnologue* (2018), the Hmong-Mien family consists of 39 languages and 9.3 million speakers. This language family has three main divisions: Hmongic, Mien, and Ho Hte. Hmongic consists of 33 languages, Mien of five languages, and Ho Hte of one (the She language). Among the linguists, one finds consensus for a putative Hmong-Mien homeland in Southern China (e.g., Benedict 1987; Kosaka 2002; Ratliff 2010; Driem 2011). The main controversary among linguists is the relationship between Hmong-Mien, Tai-Kadai, Austro-Asiatic, and Tibeto-Burman (see Ostapirat 2018 for an overview). Some linguists look for macro-family relationships, and others favor language contact.

8.2. Macro-Relationships and Hmong-Mien.

From an anthropological perspective, the temporal starting point for a discussion of the Hmong Mien language family is the arrival of rice cultivation in southeastern China during the Neolithic (e.g., Zhang and Hung 2010). A study from 2011 (Cai et al.) provides an interesting genetic perspective concerning the origins of the Hmong-Mien family. The researchers observed significant frequencies of O1b-M95, O2a-M7, and O2a-M117 among Austro-Asiatic and Hmong-Mien-speaking populations. Based on their analysis of these data, the researchers suggest a common origin language model for both language families.

An alternate interpretation of the genetic data suggests that similarities found between Hmong-Mien languages and the other languages of East Asia were shaped by the convergence of speech communities as predicted by language contact theory. Based on the data (see **Supplementary Table 16.3**), O2a-M117 mutation represents a useful marker for deciphering potential language contact between Hmong-Mien, Austro-Asiatic, Tai-Kadai and Tibeto-Burman. O1b-M95, on the other hand, helps to decipher language contact between Hmong-Mien, Austro-Asiatic, and Tai-Kadai (see **Supplementary Table 16.9**). Finally, the O2a-M7 mutation points to the language contact between Hmong-Mien and Austro-Asiatic (see **Supplementary Table 16.10**). Such a model seems consistent with more recent studies that report data for Hmong Mien speaking populations in China, Thailand, and Vietnam (Xia et al. 2019; Kutanan et al. 2020, Macholdt et al. 2020).

Martha Ratliff and her 2010 monograph provides an authoritative and comprehensive discussion of the history of Hmong-Mien from a linguistic perspective. As noted previously, she and other researchers places the putative homeland of Hmong Mien in southern China. According to Ratliff, the expansion of this language family into Thailand, Laos, and Vietnam has only occurred in the last 200 years (pp. 1-2). Additionally, her comparative analysis presents a persuasive case against a macro-family relationship with Chinese, or Tibeto-Burman, or Tai-Kadai, or Austronesian, or Austro-Asiatic. Rather, the linguistic data suggest varying degrees of lexical borrowing with these languages, with the most intensive being Chinese (pp 233-237).

8.3. Hmong-Mien and the Putative Rice Cultivators of East Asia.

The archaeological perspective, as previously discussed in Section 2.4 of this chapter, suggests that rice cultivation evolved independently in the middle Yangtze and as well as in lower Yangtze River Valley. Based on a synthesis of linguistic, archaeological, and genetic data, Hmong-Mien languages are potentially a linguistic relic of the putative rice cultivators of East Asia. Linguistic support for this argument is provided by Ratliff (2010: 231). She suggests that Hmong-Mien rice terminology has not been borrowed from another language family. Genetic support follows the strong presence of O2a-M7 and O1b-M95 mutations among contemporary Hmong-Mien speaking populations of East Asia (see **Supplementary Tables 16.9** and **16.10**) and ancient Y-chromosome data (see **Supplementary Table 16.5**) that place both mutations among the Neolithic populations of the middle Yangtze Valley. An especially strong Neolithic marker among the Hmong-Mien, the O2a-N5 mutation, which is a downstream variant of O2a-M7 (see Xia et al. 2019), provides additional support for this position.

Section 9. Koreanic.

9.1. Overview of the Linguistic Data.

Based on a discussion of archeological record, the ethnogenesis of Koreans is linked to the beginning of the Jeulmun pottery period about 10 thousand years ago (see Kim 2015). However, Kim (2009) suggests that a reliable attestation of the Korean language emerged comparatively late in their history, to about 600 years ago, when the Korean hangul script was introduced in a document called the *Hunminjeongeum*. According to the same source, classification of the Korean language has been difficult. The so-called "southern theory" attempted to associate Korean with Dravidian or Austronesian. The northern theory, on the other hand, classified Korean as part of an Altaic macro-family.

Contemporary linguistic classification of Korean has generally disassociated the language with Altaic. In their seventeenth edition from 2014, *Ethnologue* classified Korean as a language isolate. However, in the eighteenth edition, which was released in 2015, Korean was reclassified within a newly created language family called Koreanic. This language family contains just two languages, with Korean having, by far, the largest number of speakers, which totals 48 million on the Korean peninsula, and 77 million worldwide. Jejueo, the other Koreanic language, has just 5,000 speakers on Jeju Island in the Korean Straights.

9.2 Overview of the Archaeological Record.

The Paleolithic period on the Korean Peninsula is poorly documented within the archaeological record (see Kim 2015). A discussion of the Korean Neolithic begins with the origins of agriculture in China. One center of early agriculture was the Yangtze River basin and the evolution of rice cultivation (see Section 2.4). Another center of early agriculture in China is located along the Yellow River between the Mongolian steppes and Huai River (Zhao 2011). Here, roughly 8,000 years ago, the Xinglonggou culture began to cultivate foxtail and broomcorn millet. About 2,000 years later, millet cultivation expanded from this region of northeastern China to the Korean Peninsula (e.g., Stevens and Fuller 2017).

Although the arrival of early agricultural often triggered rapid population growth in many areas of the world, the situation in Korea appears to be different. The archaeological record fails to support a correlation between millet cultivation and rapid population growth during the Jeulmun period. According to Ahn (2010), millet merely supplemented a hunter-gatherer diet. Instead, it was the arrival rice cultivation from China that triggered the rapid population growth that is characteristic of early agricultural expansions, and with that, the transition from foraging to farming.

Rice cultivation came to Korea roughly 3,500 years ago during the Korean Bronze Age, or the so-called Mumun period, a term that also describes a unique form of pottery. Archaeologists (e.g., Ahn 2010) have identified three potential source regions from which this expansion may have occurred. Rice may have expanded onto the Korean Peninsula from southeastern China, somewhere near the Pearl River delta. Alternatively, rice could have expanded from central China somewhere near the Yangtze River delta. Nevertheless, most archaeologists favor northeastern China and more specifically, the Shandong and Liaodong Peninsulas. This area conveniently avoids a sea crossing and suggests that Chinese farmers migrated onto the Korean Peninsula because rice cultivation failed in Manchuria due to unfavorable climatic conditions.

9.3. Overview of the Genetic Data.

The O-M175 haplogroup represents 80 percent of the Y-chromosome variation in Korean (Kwon et al. 2015). Among the Koreans, the most informative downstream variants of O-M175 are O2a-002611, O2a-M117, O2a-F114, and O1b-SRY465. These mutations stand as genetic relics of the transition to rice agriculture on the Korean Peninsula which, as noted previously, occurred 3,500 years ago.

9.4. O2a-002611 among Koreans.

It should be noted that the O2a-002611 mutation was previously discussed in Section 3. This mutation represents a particularly strong marker of the Chinese Neolithic and the emergence of the Han ethnic group. Among contemporary Koreans, this mutation attains a frequency of around 10 percent (Kwon et al. 2015). The source of the mutation is probably eastern China.

9.5. O2a-M117 and O2a-F114 among Koreans.

At this point the reader is invited to review **Supplementary Figure 16.2**. The O2a-M134 mutation has two downstream variants, O2a-M117 and O2a-F114. Among contemporary Koreans the O2a-M117 mutation attains a frequency of around 13 percent and the O2a-F114 mutations attains a frequency around 10 percent (see **Supplementary Tables 16.3** and **16.4**). The source of O2a-F114 among Koreans is probably eastern China.

Turning now to O2a-M117, this mutation is found throughout East Asia. Unfortunately, linguistically informative downstream variants of O2a-M117 remain unknown. A possible exception is the O2a-M133 mutation, which represents almost all the O2a-M117 variation among Koreans (see Park et al. 2012). Interestingly, the O2a-M133 mutation has also been detected among the Han Chinese on Taiwan as well as among some of the aboriginal Austronesian-speaking aboriginal populations on the island (see **Supplementary Table 16.11).** As the reader may recall, the Taiwanese aboriginals are descendants of the Dapenkeng culture who migrated to the island from mainland Asia about 5,500 years ago. The Han Chinese, on the other hand, migrated to Taiwan within the last four hundred years ago (e.g., Williams 2003). Thus, the source O2a-M133 mutations among contemporary Koreans requires additional research. It might be mainland China, or it might be Taiwan. Interestingly, ancient Y-chromosome data and O2a-M133 mutations harvested from Neolithic remains along the Yellow River favor a Neolithic expansion of the mutation from China.

9.6. Koreanic and Austronesian.

Kim (2009) discussed potential Austronesian influences found in the Korean language. The presence of O2a-M133 mutation among Taiwanese aboriginals and Koreans may provide additional support for this position. Nevertheless, Kim (2009) also acknowledges that a relationship between Korean and Austronesian runs against mainstream opinion among linguists. Lee and Ramsey (2011: 27-

28) further suggest that this idea stems from Japanese linguistic research in the early twentieth century and efforts to undermine a sense of ethnic identity among the Koreans.

The genetic, linguistic, and archaeological data previously discussed in this chapter support the evolution of O2a-M117 somewhere on central plains of Yellow River. Here proto-Chinese and Proto-Tibeto-Burman separated. O2a-M117 eventually migrated westwards with speakers of early Tibeto-Burman onto the Tibetan Plateau. From Tibet, O2a-M117 then expanded southwards into India, Burma and beyond. Additionally, O2a-M117 and the O2a-M133 downstream variant expanded eastwards along the Yellow River. The eastward expansion terminated at the Yellow Sea. One population then followed the East Asian coastline northwards to Korea and another group migrated southwards to Taiwan. Such a scenario is also supported by the frequency of O2a-F114 among Han Chinese and Koreans (see **Supplementary Table 16.4**). Additionally, it seems significant that O2a-M117 and O2a-F114 are found among Mongolic and Tungusic-speaking populations (see **Supplementary Tables 16.3** and **16.4**).

This discussion of the O2a-M117, O2a-M133, and O2a-F114 mutations certainly underscores a need to further explore the downstream phylogeny of O2a-M134. Such an effort could potentially identify more informative markers that helps to decipher language variation in East Asia. Additionally, such an inquiry could help to further clarify the geographic origins and expansion of the mutation. In short, the presence of O2a-M133 among Taiwanese aboriginals and Koreans seems perplexing.

9.7. The O1b-SRY465 Mutation among Koreans.

Based on data from Kwon et al. (2015), roughly one third of Koreans have the O1b-SRY465 mutation. The same study identifies two variants of the O1b-SRY465 mutation among Koreans: O1b-47z and O1b-L682. Based on the available data, O1b-L682 appears to have evolved in Korea. O1b-SRY465 and O1b-47z, on the other hand, appear to have elsewhere as both mutations are scattered throughout East Asia (see **Supplementary Tables 16.12** and **16.13**). Kim et al. (2011) suggests that O1b-SRY465 evolved in northeastern China between 6,000 and 10 thousand years ago. The same study dates O1b-47z variants in Korea at around 4,000 years. Since O1b-SRY465 is scattered throughout East Asia, it seems as though the coastal expansion of this mutation closely follows that of O2a-M117 and O2a-F114.

9.8. Transeurasian.

The *Transeurasian hypothesis* was previously discussed in Chapters 7 and 15. According to the hypothesis, Tungusic, Mongolic, Turkic, Koreanic and Japonic all trace their origins to a common proto-Transeurasian language. An alternative view of Koreanic would view this language family as a "near isolate." As such, features shared with other language families could be explained by the convergence of different speech communities. Interestingly, the available archeological and genetic data fail to advance one model over another. Rather, interpretation of the linguistic data seems to be the decisive factor. At the end of the day, the Koreanic classification remains a highly subjective decision that entails consensus among the linguists.

Section 10. Japanese and Haplogroup O-M175.

10.1. Introduction.

The internal linguistic phylogeny of the Japonic family, as well as the geographic distribution of Japonic language-speakers, seems to suggest that Japonic behaves much like a "near isolate" like Koreanic. Japonic has two main branches, Japanese with a single language, and Ryukyuan with eleven languages. The Japanese branch has, by far, the largest number of speakers, around 129 million. While the number of Ryukyuan speakers is unknown (e.g., Shimoji 2010), data from *Ethnologue* (2018) suggests around 2,000 speakers. In terms of geographic distribution, Japanese is spoken throughout the entire range of the Japanese islands. Ryukyuan languages, on the other hand, are confined to the Ryukyuan Islands at the southernmost tip of the Japanese archipelago.

10.2. Linguistic, Anthropological, and Genetic Perspectives.

In terms of geography, contemporary Japan consists of a chain of islands that extend roughly 3,000 kilometers north to south. As mentioned previously in Chapters 4 and 6, the gene pool of contemporary Japan has a strong Paleolithic component, roughly 40 percent. The genetic relics of the Paleolithic founding populations of Japan are the C1a-M8 and D1b-M55 mutations. The genetic data are consistent with the archeological record which suggests that modern humans colonized the Japanese Islands roughly 30 thousand years. Around 16 thousand years ago their descendants evolved into the Jomon hunter-gatherer culture (e.g., Hudson 2013). The term "Jomon" describes a unique style of pottery that has become a signature relic of these people.

Roughly 2,000 years ago the Yayoi culture migrated from Korea to the Japanese island of Kyushu. They introduced rice agriculture which eventually replaced foraging as the main subsistence strategy. The O1b-SRY465 and O1b-47z mutations are the genetic relics of the Yayoi migration (e.g., Hammer et al. 2006). According to Naitoh et al. (2013) and Sato et al. (2014), roughly a third of Japanese have O1b-SRY465 mutation and one of its downstream variants. As such, the genetic evidence potentially supports linguistic arguments (e.g., Whitman 2012) that posit a common ancestral language for Koreanic and Japonic. Nevertheless, some linguists question this relationship (e.g., Tranter 2012). Rather, Japonic and Koreanic are essentially seen as language isolates that influenced each other as the result of language contact and the expansion of agriculture from the Korean Peninsula to the Japanese islands.

Efforts to build a linguistic macro-family relationship for Japonic and Koreanic are controversial. However, language contact influence exerted by Chinese in both language families is incontrovertible. For example, roughly half the lexicon in both language families has a Chinese origin (Kim 2009; Shibatani 2009). The influence of Chinese extends, in fact, to the earliest attestations of Koreanic and Japonic and efforts to adapt Chinese character script for writing Old Korean and Old Japanese (e.g., Tranter 2012). Given the indisputable influence that Chinese has played in shaping the Korean and Japanese languages, it should not be surprising to detect the genetic signature of Han Chinese in the contemporary Japanese and Korean gene pools. Indeed, roughly 10 percent of Koreans and 5 percent of Japanese have the O2a-002611 Chinese signature mutation (see **Supplementary Table 16.2**).

Building a macro-relationship for Japonic and Koreanic seems problematic because it ignores massive Chinese language influence in both families and the potential contribution to Japonic from the Jomon hunter-gatherers. Like Koreanic, an alternative view of Japonic would certainly view this language family as a "near isolate." Like Koreanic, the available archeological and genetic data fail to advance a macro-family model over a language isolate model for Japonic. Once again, interpretation of the linguistic data seems to be the decisive factor.

10.3. Japanese and Austronesian.

According to Naitoh et al. (2013) and Sato et al. (2014), about 8 percent of Japanese have the O2a-M134 mutation. As previously detailed, better resolution of O2a-M134 variation among Koreans may clarify the extent of Austronesian language contact with speakers of early-Koreanic languages. Taking this a step further, clarification of O2a-M134 mutation among Japanese may also clarify the extent of Austronesian language contact with speakers of early Japonic. Interestingly, Robbeets (2017b) presents linguistic arguments for language contact between early Japonic and early Austronesian speech communities. The researcher favors a "para-Austronesian" speech community on the Shandong peninsula of northeastern Chinese mainland. She also places the putative homeland of Japonic languages on the Chinese mainland and the Liaodong peninsulas. According to the researcher, both speech communities converged roughly 4,000 years ago.

Another possible model for explaining potential Austronesian and Japanese would posit an Austronesian expansion from Taiwan onto the southern Ryukyuan islands beginning about 4.500 years ago. Archeological support for the Ryukyuan model comes from tool and pottery remains found on the Japanese Yaeyama Islands, which are located at the southernmost tip of Ryukyuan atoll, about 250 kilometers east of Taiwan. Two reports (Summerhayes and Anderson 2009; Hudson 2017b) identify these items as possible Austronesian artifacts from Taiwan. Additionally, both reports raise the possibility of contact between Jomon hunter-gatherers and Austronesians in the vicinity of Okinawa Island. While the available archeological evidence fails to support an Austronesian migration onto Okinawa or points beyond, better resolution of O2a-M117 and O2a-N6 variation among East Asians may well paint a different story. Alternatively, trade relationships in the Taiwan Straits and along the Ryukyuan Islands may have resulted in the convergence of speech communities.

Figure 16.2. Korea, Japan, Kyushu, Okinawa, Ryukyu Islands, Taiwan, and China.

Section 11. Conclusions for Haplogroup O-M175.

The contemporary distribution of O-M175 mutations stand as a genetic relic of the Neolithic transition in East Asia, South Asia, Island Southeast Asia, and Oceania. Among the linguistic relics of this transformation is Chinese, Tibeto-Burman, Austro-Asiatic, Austronesian, Koreanic, Japonic, Tai-Kadai, and Hmong-Mien. In East Asia, rice agriculture fueled an expansion of Chinese, Austro-Asiatic, and Austronesian languages. In the same region a different evolutionary trajectory is observed for Koreanic, Japonic, Tai-Kadai, and Hmong-Mien. With these languages the data suggest *in situ* co-evolution of rice agriculture and language. The evolutionary history of Tibeto-Burman languages is somewhat unique in that the initial expansion of this language was fueled by barley rather than rice agriculture. Finally, the expansion of Austronesian was initially fueled by rice agriculture and later transitioned into the cultivation of tubers and tree crops in the lowland coastal regions of Island Southeast Asia.

Rice paddy agriculture is ideally suited to the climatic conditions of South and East Asia, and in particular, areas that receive the annual monsoon rains. As such, this crop supports the high population density that is characteristic of both regions. Taking this a step further, this explains why East Asian languages now occupy a huge corner of the tapestry of language variation. Chinese, for example, is the first language of over a billion people in the region. Similarly, barley cultivation is ideal for high altitude regions, such as the Tibetan Plateau. This explains the position attained by Tibeto-Burman with the global tapestry of language variation. Finally, the cultivation of tubers and tree crops elucidate the expansion of Austronesian languages. Unlike grains, tubers and tree crops thrive and survive in the tropics.

The presentation of data in this chapter also invites researcher to focus not only cultural adaptation that have influenced language variation, but also evolutionary adaptations. The success of the Austronesian language family was facilitated by resistance to tropical splenomegaly syndrome, a condition caused by chronic exposure to malaria. Similarly, evolutionary responses to hypoxia explain the success of Tibeto-Burman languages.

Chapter 17: Haplogroup Q-M242.

Section 1. The Contemporary Distribution of Haplogroup Q-M242.

Although the Q-M242 haplogroup evolved in Asia, its downstream mutations have become an important tool for deciphering indigenous language variation in the Americas. The reader is now directed to **Supplementary Table 17.1.**, which reports contemporary data. As shown by the table, the Q-M242 mutation is distributed throughout Eurasia, where it generally attains a small percentage when detected within a population. The reader should now review **Supplementary Tables 17.3** and **17.4** which reports Q-M242 data for contemporary Native Americans. As shown by the tables, the Q-M242 haplogroup represent almost all the indigenous genetic diversity.

The term "indigenous genetic diversity" requires an explanation. Historical contact with Europeans, which occurred after 1492, explains the substantial presence of West Eurasian Y-chromosome haplogroups, such as R-M207, among populations that identify themselves as Native American. As such, deciphering the prehistory of the indigenous languages of the Americas entails the identification of Native American genetic lineages that predate the arrival of Christopher Columbus. Downstream variants of haplogroups Q-M242 and C2-M217 are the "indigenous" Y-chromosome mutations. In North America, Q-M242 mutations represent about 93 percent of indigenous Y-chromosome diversity (e.g., Zegura et al. 2004). Here, the remaining 7 percent of Native American genes belong to C2b-P39. In Central and South America, on the other hand, Q-M242 represents almost all of the indigenous Y-chromosome diversity (e.g., Geppert et al. 2011; Roewer et al. 2013).

The search for indigenous Y-chromosome mutations among Native American extends back to 1996 when Underhill et al. reported the discovery of the Q1b-M3 mutation. Despite this early success, the search for informative Native American mutations has remained, elusive. Much of the data has only been recently published (e.g., Grugni et al. 2019). Part of the problem could stem from post-Columbian factors that reduced Y-chromosome variation among Native Americans. This conforms to a population model called bottleneck effect. Shortly after the arrival of the Europeans in the fifteenth century, many Native Americans groups experienced a rapid decrease in population size. Many of them succumbed to European diseases such as smallpox, against which they had not developed immunity. Furthermore, admixture between European men and Native American women may have reduced Y-chromosome genetic variation among the indigenous populations of the New World (for additional details, see Malhi et al. 2008).

Pre-Columbian factors may also explain the limited genetic variation found in Native Americans. Humans colonized the Americas relatively late in the game, about 15 thousand years ago. Thus, genetic variation may be a question of time depth. Populations in Africa, for example, have diversified for around 300 thousand years. A demographic model that surfaced in two studies (Regueiro et al. 2013; Roewer et al. 2013) may also explain factors that reduced genetic variation among Native Americans prior to the arrival of Columbus. Both studies suggest that populations in Africa and Eurasia have a significantly different demographic history than populations in the Americas. Compared to Africa and Eurasia, the Americas never experienced a massive expansion of genetic variation that is characteristic of agriculture expansions. Battaglia et al. (2013) suggest, for example, that the reduced variation observed in indigenous Americans may reflect a preference for hunter-gathering over agriculture as the primary survival strategy, even in areas where crops were cultivated. Bellwood (2005:

146-149) suggest that agriculture was limited in the Americans because of climate, the limited number of animals available for domestication, and because the only cereal crop was maize. He also questions the extent to which agriculture was a significant part of the survival strategy among those who cultivated crops. For example, maize, potatoes, and manioc are a potential source of calories. However, condiment crops, such as chilies and avocados, and the cultivation of squash for drinking gourds, are not staples.

Interestingly, language typology seems to reflect an inverse correlation between linguistic and genetic variation in the Americas. In other words, the characteristic leveling of linguistic diversity that accompanied agriculture expansion in the Old World may not have occurred in the New World. Compared to Africa, Eurasia, and Oceania, linguistic diversity in the Americas appears much more diverse and more difficult to classify. *Ethnologue* (2016) lists a total of 82 language isolates for the world, and of these languages, 60 are found in the Americas. Furthermore, of the 62 unclassified languages listed by *Ethnologue* (2016), 30 are found in the Western Hemisphere, and more specifically, in South America (for more information, see Supplementary Table 17.2).

Despite the difficulties listed above, it is important to emphasize that the effort to identify informative Y-chromosome mutations among Native Americans is finally gaining momentum, especially in the last five years. However, more data are needed, especially for North America. Unfortunately, many of the North American native groups refuse to participate in genetic studies because of their historical mistrust of Europeans (e.g., Reardon 2017). Mulligan and Szathmary (2017) also suggest that Native Americans feel disrespected by researchers, and for this reason, they refuse to participate in genetic studies. According to their paper, some researchers strive to identify an Asian source population for Native Americans. However, contrary to what is reported by these researchers, some Native Americans take the position that they came from the Americas and to suggest otherwise is offensive.

Section 2. Upper Paleolithic Mammoth Hunters and Q1b-M346 Mutations.

2.1. Introduction.

A synthesis of multidisciplinary perspectives suggests that the prehistory of the Americas and Eurasia are linked by the Upper Paleolithic mammoth hunter tradition. The genetic relic of this tradition is the Q1b-L330 mutation in Siberia, and the Q1b-M3 and Q1b-Z780 mutation among Native Americans. See **Supplemental Figure 17.1**).

2.2. Phylogenetic Relationships within Q-M242.

The reader is invited to locate the KR-M526 paragroup in **Supplementary Figure 1.1** from the first chapter. As previously detailed in Chapter 13, the P1-M45 mutation, which is downstream from KR-M526, stands as a genetic relic of human expansions across the North Eurasian landmass during Marine Isotope Stage 3. According to Poznik et al. (2016), the P1-M45 mutation evolved roughly 47 thousand years ago. Diverging from P1-M45 are Haplogroups Q-M242 and R-M207. The Q-M242 mutation evolved roughly 33 thousand years ago.

At this point the reader is invited to examine **Supplementary Figure 17.1**. The Q-M242 haplogroup has two main internal divisions within its phylogeny: Q1-F903 and Q2-L275. Both mutations evolved about 30 thousand years (Poznik et al. 2016). Contemporary Q2-275 mutations are confined to Eurasia and represent a very small fraction of the genetic diversity within this region (e.g., Huang et al. 2018). Furthermore, Q2-L275 mutations are not linguistically informative. Rather, linguistically informative mutations are downstream from Q1-F903.

Turning now to the diversification of Q1-F903 in Asia, as shown by **Supplementary Figure 17.1**, downstream from the Q1-F903 mutation are the Q1a-F1096 and Q1b-M346 mutations. Downstream from Q1b-M346 is the Q1b-L54 mutation. According to Wei et al. (2018), this mutation evolved roughly 17 thousand years ago in south-central Siberia. Three linguistically informative lineages evolve from Q1b-L54: Q1b-L330, Q1b-M3, and Q1b-Z780. Q1b-L330 represents an important mutation among the contemporary Ket people of southern Siberia. The Q1b-M3 and Q1b-Z781 mutations evolved roughly 14.5 thousand years ago and represent the beginning of *in situ* genetic diversification of the Q-M242 haplogroup in the Americas (Wei et al. 2018a).

2.3. Marine Isotope Stage 3 Expansions in Northern Eurasia.

Mammuthus primigenius, commonly known as the woolly mammoth, evolved roughly 450 thousand years ago. During the Last Ice Age, they occupied a vast territory from Europe to the Americas (Kahlke 2015). At the beginning of Marine Isotope Stage 3, roughly 50 thousand years ago, *Homo sapiens* expanded across Northern Eurasia. During this expansion human populations survived by hunting the woolly mammoths that proliferated in this region. Archaeological support comes from the Sopochnaya Karga meteorological station which is located above the Arctic Circle in Siberia. Near the station researchers discovered the remains of a woolly mammoth that died about 45 thousand years ago. Examination of the remains indicates that humans killed and butchered the animal (Pitulko et al. 2016).

Additional archeological evidence for mammoth hunting comes from the Sunghir archeological site located about 190 km northeast of Moscow. The remains of five males were found. They died roughly 34 thousand years ago. Evidence from the site further suggests that they hunted mammoths. One of the remains, the so-called Sunghir-1 man, was between 35 and 45 years old at the time of his death, which may have been the result of a hunting accident. According to speculation, the spear hit him instead of the mammoth. Moreover, those who buried the man appear to have conducted a funeral ritual. He was buried with valuable stone tools. Thousands of mammoth ivory beads were also sown onto his burial garments. Finally, his corpse was also covered in red ochre (for additional information, see Sikora et al. 2017). Important archaeological evidence also comes from the Yana Rhinoceros Horn site located above the Arctic Circle, where the Yana River empties into the Arctic Ocean. Archeological remains suggest that the site was used by Paleolithic mammoth hunters (Pitulko et al. 2004; Nikolskiy and Pitulko 2013). Dental remains were recovered from two boys who died here roughly 32 thousand years ago. Researchers determined that they have the P1-M45 mutation (Sikora et al. 2019).

As suggested by the above archeological data, Upper Paleolithic foragers successfully adapted to the cold climate of Siberia during Marine Isotope Stage 3. Part of this adaptation required specialized hunting skills needed to harvest a woolly mammoth, an animal that is about the same size as a modern-day African elephant. A gruesome discussion of the tactics utilized by Paleolithic hunters is provided by Pitulko et al. (2016) and one thing seems obvious - it must have been a dangerous undertaking. Nevertheless, the reward must have outweighed the risk. Successful prehistoric human adaptation to cold climate required a reliable high energy food supply, adequate clothing and shelter, raw materials for making cutting tools and projectile points, and fuel for fire. The woolly mammoth solved all these problems (Pitulko and Nikolskiy 2012; Pitulko et al. 2016; Pfeifer et al. 2019). A single woolly mammoth provided thousands of kilograms of meat which could be cached and stored long term in sub-zero conditions. Mammoth ivory produced exceptional projectile points. The hide was used for clothing and tents. Dung and bones became a source fuel for a fire. Of course, cold adapted Paleolithic people in northern Eurasia ate other animals. Nevertheless, mammoth remains provide especially robust archaeological evidence that points to cultural continuity that had begun roughly 50 thousand years ago in Northern Eurasia and terminated both in this region and in the Americas roughly 11 thousand years ago with the onset of the Holocene.

2.4 Marine Isotope Stage 2 Hiatus.

The Last Glacial Maximum occurred about 27 thousand years ago (Clark et al. 2009). At this stage of the Last Ice Age, glaciation had reached its southern most extent across the Northern Hemisphere. While glaciation in Siberia was not as extreme as that in Europe, much of the region was nevertheless cold, arid, and uninhabitable (e.g., Serdyuk 2005). Most of the region became depopulated. Foragers who once hunted above the Arctic Circle retreated below the 50th parallel. This depopulation of Northern Eurasia is characterized by Hamilton and Buchanan (2010) as a "hiatus" that lasted between 32 thousand and 16 thousand years ago. A simple explanation of "hiatus" entails the movement of hunter-gatherers into refugia where they waited for better weather.

Ice Age hiatus altered the Y-chromosome landscape of Northern Eurasia because the refugia consisted of reproductively isolating populations of humans (see, for example, Stewart and Stringer 2012; Gavashelishvili and Tarkhnishvili 2016). The genetic relics of hiatus include haplogroups Q-M242 and R-M207. Archeological and genetic support for this conclusion comes from the banks of the Belaya River, near Lake Baikal, in southern Siberia and the remains of the Mal'ta boy. Raghavan et al. (2014) report that he died about 24 thousand years ago, roughly at the time of the Last Glacial Maximum. The researchers further report Y-chromosome data that place the child somewhere near the root of the R-M207 haplogroup, where haplogroups R-M207 and Q-M242 diverge from P1-M45.

2.5. Marine Isotope Stage 2 and the Mammoth Hunters.

The Altai-Sayan region of southern Siberia was a refugia that protected hunters-gatherers from the extreme glaciation and adverse weather conditions characteristic of the Last Glacial Maximum (Binney et al. 2017). Here, the contemporary borders of Russia, Kazakhstan, Mongolia, and China converge on a map. Several Y-chromosome studies (e.g., Zegura et al. 2004; Dulik et al. 2012; Wei et al 2018) identify this region as the geographic point of origin for Upper Paleolithic human expansions into North America. Strong Y-chromosome support for this position stems from contemporary data and the high frequency of the Q1b-L330 mutation among the Ket (see Flegontov et al. 2016b). They are among the indigenous peoples of the Altai-Sayan region and speak a Yeniseian language. Archaeological support comes from numerous archaeological sites within the region, many of which are located in the Yenisei River Valley. Graf and her 2009 analysis of these archaeological data reports human occupation of the Yenisei River Valley by around 33 thousand years ago. The subsistence strategy of the Upper Paleolithic people who lived here included mammoth hunting.

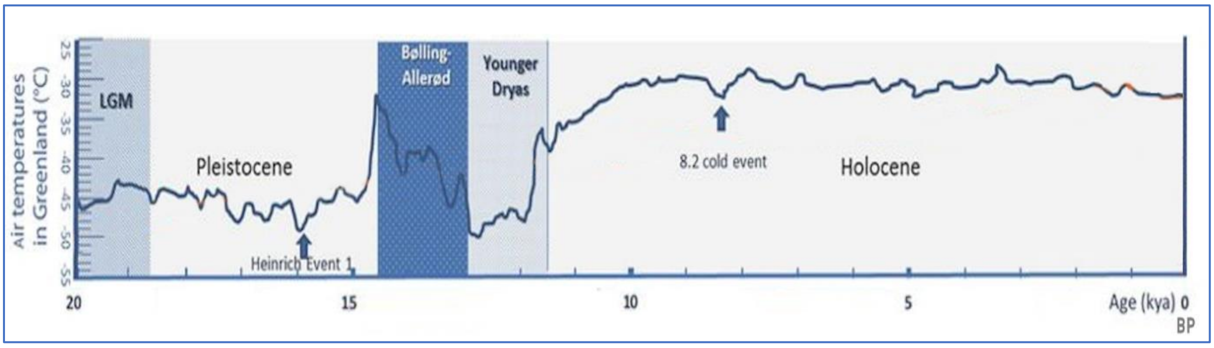
Figure 17.1. Altai and Sayan Mountains, Yenisei River, and Lake Baikal.

Hamilton and Buchanan (2010) time the end of the Marine Isotope Stage 2 hiatus in the Altai-Sayan region at around 16 thousand years ago. Weather conditions in southern Siberia consisted of widely oscillating warm and cool phases (Serdyuk 2005). As a result, woolly mammoth populations in this region contracted and expanded rapidly based on the availability of forage, which was mediated by climatic conditions (Pitulko and Nikolskiy 2012; Mann et al. 2015). A brief period of favorable weather, the Bølling-Allerød warming phase, and the corresponding increase in forage, explain why mammoth hunters and Q1b-L54 mutation expanded from the Altai region into the Americas roughly 16 thousand years ago. According to Pitulko and Nikolskiy (2012), the sudden increase in forage produced a sudden increase in mammoth populations. Population pressure then forced the mammoths to migrate across northeastern Siberia into Alaska in search of more forage. Taking this a step further, when the mammoths thrived, the hunters thrived; and when the mammoths migrated, the hunters followed.

The above model of the Upper Paleolithic expansion into the Americas is supported by the Y-chromosome data. Wei et al. (2018a) time the evolution of Q1-L54 to a gradual increase in air temperatures that preceded the Bølling-Allerød warming phase. The study further suggests that southern Siberian origins of Q1b-L54 are supported by the current distribution of its downstream variants, Q1b-L330, Q1b-M3 and Q1b-Z781. Q1b-L330 remained in southern Siberia and as previously mentioned, became a predominant lineage among the Ket people of the region. Q1b-M3 and Q1b-Z781, on the other hand, evolved roughly 15 thousand years ago and mark the beginning of Y-chromosome diversity in the New World.

Additionally, dating estimates provided by Wei et al. (2018a) for Q1b-M3 and Q1b-Z781 correlate well with archaeological data from Swan Point, which is in the Tanana Valley about 100 km southeast of Fairbanks, Alaska (Holmes 2011). Based on carbon-14 data obtained from carbonized grease and fat remains, humans occupied this site roughly 14 thousand years ago. Bones suggest mammoth and horses were on the menu. Additionally, evidence suggests that people used bones as fuel, a common practice among the Upper Paleolithic mammoth hunters of Siberia. Finally, their tools are similar to those manufactured by the Upper Paleolithic people of northeastern Asia

Figure 17.2. Bølling-Allerød Interstadial. Source: Wikipedia and Platt, D. et al. 2017.



2.6. The Mammoth Hunter Tradition in North America.

The oscillating pattern of cooling and warming periods during the Late Pleistocene ended with the Younger Dryas, a brief cold snap that lasted about 800 years, between 12.9 and 11.7 thousand years ago. The end of the Younger Dryas marks the beginning of the Holocene, and with that, higher temperatures and more stable weather conditions. Warmer weather then caused the ice glaciers to melt, which caused sea levels to rise. As the sea level rose, the Bering land bridge, a corridor that once connected Russia and Alaska, eventually disappeared.

As previously noted, the gradual geographic isolation of New World populations from those in Asia is recorded on the human Y-chromosome by downstream variants of the Q1b-L54 mutation. For example, Anzick-1, the remains of a boy who died about 13 thousand years ago in Western Montana, has a mutation that is downstream from Q1b-Z780. Shuka Kaa, the remains of a man who died about 10 thousand years ago on Prince of Wales Island in Alaska, has the Q1b-M3 mutation (see Kemp et al. 2007).

Returning to the anthropological perspective, after the mammoth hunters reached eastern Alaska about 15 thousand years ago, the Cordilleran and Laurentide ice sheets temporarily blocked further migration into contemporary British Columbia. The process of deglaciation eventually produced an ice-free corridor that allowed the hunters to continue the journey onto the Great Plains of North America (Dyke 2004; Potter et al. 2018). The archaeological record suggests that the journey through the ice-free corridor occurred by around 14,000 years ago. This figure is derived from dating estimates taken from a North American mastodon, a proboscidean closely related to the mammoth. The remains were uncovered at the Manis archaeological site in Washington State (Waters 2011). Researchers determined that the mastodon was killed by humans because a bone or antler projectile was found imbedded in one of the rib bones. Additional evidence for timing the passage through the ice-free corridor comes from dating estimates for the Anzick-1 child that found at a burial site in Western Montana. As noted previously, he died about 13 thousand years ago.

The Anzick boy represents a significant find because the artefacts found at his grave site are clearly linked with the so-called Clovis culture and their use of a unique projectile point. Traditionally, the Clovis culture is interpreted as the initial Native American cultural tradition (see Williams and Madsen 2020 for more details and a dissenting opinion). An alternative interpretation, one that linguists should consider, is that Clovis points merely represents the end of an Upper Paleolithic mammoth hunting tradition that began about 50 thousand years ago in Northern Eurasia. Frison (1998) suggests that mammoth hunting required a robust projectile point and Clovis points were very much up to the task. As such, Clovis points were a Paleolithic innovation in North America that sustained a subsistence strategy that had begun tens of thousands of years earlier in Northern Eurasia. Cultural continuity between the mammoth hunters of Eurasia and those of North America is supported by similar burial rituals. Sunghir-1, one of the first Upper Paleolithic mammoth hunters, and Anzick-1, one of the last Upper Paleolithic mammoth hunters, were covered in red ochre and buried with valuable tools that were needed for hunting large herbivores (cf. Rasmussen et al. 2014; Sikora et al. 2017).

According to Frison (1998) mastodon and mammoth hunting continued in North America until about 11 thousand years ago. Then, the mammoths suddenly disappeared. Perhaps they became extinct because of warmer weather, or perhaps it was human population pressure, or perhaps a combination of both (Fiedel 2008). Regardless of the reasons, the demise of mammoth hunting potentially marks the beginning of the Native American cultural tradition. Like in the Americas, the mammoths of Eurasia also disappeared at the beginning of the Holocene (e.g., Orlova et al. 2001). Like in the Americas, the demise of the Eurasian mammoths is linked to climate change and over-hunting (Nogués-Bravo et al. 2008). From a Y-chromosome, the demise of the mammoth hunter tradition explains the contemporary distribution of haplogroup Q-M242 mutations in the Americas. In Eurasia, the demise of the same tradition explains the contemporary distribution of haplogroup R-M207 mutations (see Chapter 18).

Section 3. A Working Model of Native American Origins for Linguists.

As previously noted, genetic studies have focused on defining a source population for Native Americans and conclude that they came from Asia. The archaeologists have traditionally associated the Clovis tool making tradition with the first Native American cultures. For the purposes of linguistic research, it might be preferable to define the term "Native American" as a cultural tradition that evolved in the Americas. Such a position seems more consistent with recent archaeological, climatological, genetic, and linguistic data that are presented in this chapter.

A good working model of Native American origins begins with the mammoth hunter cultural tradition that evolved in northern Eurasia roughly 50 thousand years ago. By 14 thousand years ago, this cultural tradition of the Upper Paleolithic had expanded onto the Great Plains of North America. With the arrival of the Holocene, about 11 thousand years ago, this cultural tradition, which once spanned the entire northern hemisphere, suddenly disappeared. The mammoths, mastodons, and other mega-herbivores could not adapt to the change in climate. The humans could. In the Americas, the former mammoth hunter cultural tradition of the Upper Paleolithic came to an end and a new Native American cultural tradition evolved. This new cultural reflects adaptation to climate change and ultimately substantial changes in subsistence strategy.

Section 4. Classification of Native American Languages.

The above model of Native American origins provides a time component for building models of indigenous language variation in the Western Hemisphere. Now comes the question of information management or how the data should be organized to construct these models. A tripartite division of the data into Amerind, Eyak-Athabaskan, and Eskimo-Aleut seems to work. This division follows Greenberg (1987) and his view on linguistic diversity in the Americas. Since Greenberg's classification is controversial among the linguists, an explanation is in order.

American Indian Languages, a reference guide published in 1997 by Lyle Campbell, is an authoritative must-have resource for linguists. The guide examines Native American languages from the perspective of historical linguists with the goal of classifying these languages. According to Campbell, the methodology employed by Greenberg is unconventional and unreliable. Campbell asserts that a rigorous application of the comparative method fails to prove the unity of Amerind as suggested by Greenberg.

Campbell's classification of Native Americans languages is consistent with contemporary mainstream opinion (e.g., *Ethnologue* or *Glottolog*). Nevertheless, in the last thirty years Greenberg's work has often surfaced in genetic studies as an authoritative classification for Native American languages. So frequent was the use of Greenberg's classification that Bolnick et al. (2004) felt compelled to warn geneticists. In an article published in a science journal they labeled Greenberg's classification as problematic and non-standard. Nevertheless, Greenberg still surfaces in the genetic studies (e.g., Flegontov et al. 2016a).

Campbell's and Greenberg's approach to classification illustrates a strange dichotomy between a single discipline historical linguistic approach to Native American languages and one that integrates multidisciplinary perspectives. The historical linguistic approach clearly sides with Campbell. However, Greenberg provides a natural division of the data when multidisciplinary perspectives are employed. From a genetics perspective, Amerinds represent *in situ* diversification of Q1b-L54 mutations in the Western Hemisphere. The genetic history of Eskimo-Aleut and Eyak-Athabaskan speakers, on the other hand, is potentially shaped by later geneflow across the Bering Sea. Culturally, Amerind reflects *in situ* adaptation to climate change in the New World. Eskimo-Aleut culture was hugely influenced by the development of active whaling which began in Japan roughly 5,000 years ago. Linguistically, Amerind represents *in situ* diversification of languages in the Western Hemisphere since the beginning of the Holocene. Athabaskan and Eskimo-Aleut languages were potentially shaped by later cultural exchange with the indigenous peoples of northeastern Asia.

The adoption of Greenberg's classification to subdivide the Native American data should not be interpreted as a disagreement with the contemporary standard classification of Native American languages (e.g, Campbell 1997; or *Ethnologue*; or *Glottolog*). Rather, the subdivision of the data into

Amerind, Eyak-Athabaskan and Eskimo-Aleut achieves a more elegant presentation of the Native American linguistic diversity when multidisciplinary perspective are employed.

Section 5. Amerinds.

5.1. Overview.

This section divides the Amerind data into five regional groups: Pacific Coast, Eastern Woodlands, American Southwest, Central America, Central Andes Mountains, and Amazonia. A synthesis of the data for these groups indicates the following: the pattern and incredible diversity of Amerind languages reflects the diversified and regionally influenced subsistence strategies of the pre-Columbian Amerind cultures.

Coinciding with the initial diffusion of Q1b-M3 and Q1b-Z780 mutations in the Americas was the beginning of the Holocene about 11 thousand years ago. The Bering land bridge became submerge under the rising sea level. The mammoths and other mega faunal resources disappeared. Former mammoth hunters now exploited new food resources found in the regions where they had settled. The earliest example of this transition comes from the Upper Sun River site in Central Alaska. Human remains and artifacts found at this location are dated to about 12 thousand years ago. Here, researchers found evidence of a more diversified diet that included salmon (Potter et al. 2014).

The term "Amerind" represents about a thousand indigenous languages in the New World (see Supplementary Table 17.2). Thus, it goes without saying that this monograph could not possibly provide a comprehensive discussion of this incredible diversity of Amerind language and culture. Rather, the goal is to find examples that will help model the prehistory Amerind languages from a triangulated Y-chromosome-based perspective.

Y-chromosome data for contemporary Amerind populations is limited in terms of resolution. From the published population reports, researchers basically have frequency data for the Q1b-M3 and Q1b-Z780 mutations. Fortunately, ancient DNA from human remains helps to fill the gaps. The available ancient DNA data suggest that the human colonization of the Americas, from Alaska to Patagonia, occurred within perhaps a few thousand years. Given the fact that both regions are separated by a distance of 15 thousand kilometers, the pace of human expansion throughout the Americas occurred fairly rapidly. Evidence for this conclusion comes from the Shuka Kaa remains in Alaska and Sumidouro Cavern remains from Brazil (see Supplementary Table 17.6). Both sets of remains belong to the Q1b-M3 mutation. The dating results from both sites are strikingly similar, about 10 thousand years ago.

5.2. Pacific Coast Indians.

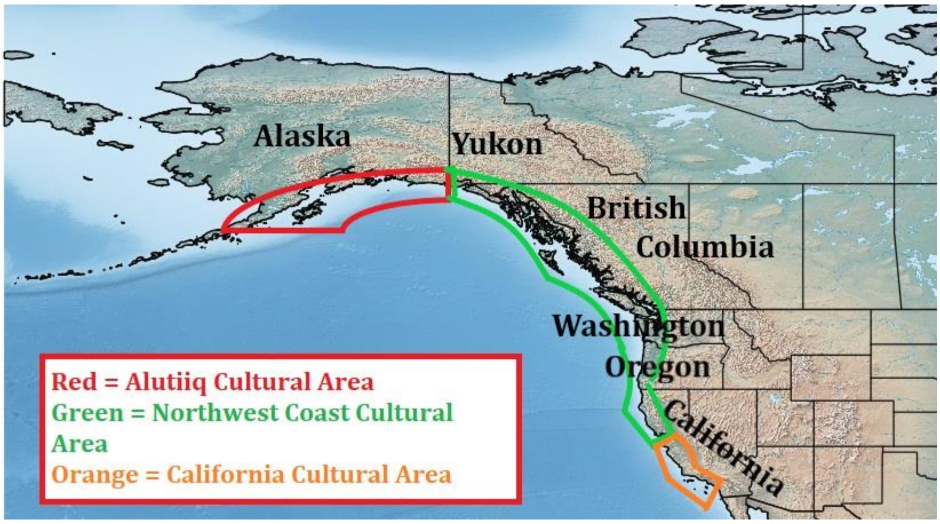
As discussed previously, about 15 thousand years ago mammoth hunters crossed over the Bering land bridge from northeastern Asia to Alaska. The mammoth hunters then expanded southwards onto the Great Plains of North America. According to archeological and climatological data, this was facilitated by an ice-free corridor between the Cordilleran and Laurentide ice sheets. Other models, however, favor a second southward migration along the Pacific coastline of the United States and Canada. A recent report (Potter et al. 2018) suggests that although a coastal migration remains plausible, such a model is not supported by the archeological record. Moss and Erlandson in their 1995 paper discuss the terrain along the North American coastline and suggest that mountains, tectonic activity, and the lack of a coastal plain would have hindered a coastal migration. Rather, as suggested by Erlandson, Moss and Des Lauriers (2008) the settlement of the Pacific Coast began with migrations from the North American interior. These migrations then spread westwards alongside rivers that empty into the Pacific Ocean.

The indigenous peoples who eventually settled along the Pacific coastline of North America lived in relatively permanent settlements. According to Moss and Erlandson (1995), these settlements exhibit high population density that can be attributed to an abundance of marine resources such as sea mammals and shellfish. The same report divides the indigenous peoples of the North America Pacific coast into three cultural areas: the Alutiiq cultural area, the northwest coast cultural area, and the California cultural area. The Alutiiq cultural area runs along the coastline of southern Alaska which is the home of the Alutiiq people. They are sometimes referred to as Pacific Eskimos or Pacific Yupik. Their language belongs to the Eskimo-Aleut family.

The northwest coast cultural area begins in southwestern Alaska, where Alaska, the Yukon Territories and British Columbia converge. It terminates in northern California near San Francisco. Linguistic diversity within this area is complex. Here, the Eyak-Athabaskan people and languages have an interesting "leapfrog" distribution. The Tlingit, for example, are Eyak-Athabaskan people found at the northernmost part of the cultural area in southwestern Alaska. At the southernmost part, in California and Oregon, are the Pacific Coast Eyak-Athabaskan languages and peoples. In addition to Eyak-Athabaskan languages, linguistic diversity along the northwest coast cultural area includes the Haida language family. Additionally, cultural diversity includes Tsimshian, Wakashan, Salish, Chimakuan languages and peoples. Finally, linguistic diversity includes Ritwan, a sub-branch of the Algic language family. This seems expected because Algic (or Algonquin) is a major indigenous language family of the eastern United States and southern Canada.

The California cultural area runs south of San Francisco and includes the Channel Islands near Santa Barbara. For linguists, the Channel Islands and surrounding region are significant because of its historical association with the Chumash people and languages.

Figure 17.3. Indigenous Cultures along the North America Pacific Coast. Data Source: Moss and Erlandson (1995).



Turning now to the genetic data, contemporary Y-chromosome population data for Pacific coast Indians are limited to the Tlingit and Haida (see **Supplementary Table 17.3**, **17.4**, and **17.5**). However, we have ancient DNA data for three remains from the California cultural area (**Supplementary Table** 17.6). The St. Miguel Island and Point Sal remains belong to Q1b-M924. Remains from San Nicholas Island belong to Q1b-Y4276. The Q1b-M924 mutation potentially connects the Chumash with Amerinds and an overall model of *in situ* cultural diversification since the onset of the Holocene. The Q1b-Y4276 mutation, on the other hand, may connect the California cultural area with Algic languages. Ritwan, and its attested presence in northern California, provides linguistic support for this argument. Genetic support stems from Grugni et al. (2019). This study identifies Q1b-Y4276 as a potential marker for Algic peoples of the northeastern United States. Alternatively, Q1b-Y4276 may link the California cultural area with Eyak-Athabaskan. A downstream variant of Q1b-Y4276, the Q1b-B34 mutation, was found in ancient remains at an Athabaskan cemetery near Kenai, Alaska (**Supplementary Table 17.6**).

5.3. Eastern Woodlands.

The Eastern Woodlands cultural area is found in the United States and east of the Mississippi River, as well as southern Ontario and Quebec in Canada. From an archaeological perspective, this region was inhabited around the beginning of the Holocene (Snow 2013: 354). Unlike the indigenous peoples of the Pacific coast of North America, the indigenous peoples of the Eastern Woodlands supplemented their subsistence strategy with farming. Farming within the Eastern Woodlands may have fueled an expansion of the Algic, Iroquoian, Siouan-Catawban, and Muskogean language families (Bellwood 2005: 174-179). Crop cultivated in this region by the Native Americans included maize, beans, squash, sunflower, tobacco, and goosefoot.

The Algic language family is often called "Algonquin." Campbell (1997:156) places the putative homeland of Algic languages around the Great Lakes but notes that some have placed it further west. Algic languages are distributed over a vast area. As previously mentioned, this language family is found in the eastern United States and northern California. Moreover, Algic languages are distributed across much of southern Canada. Finally, some of the Algic-speaking peoples, such as the Cheyenne and Arapahoe, inhabited the Great Plains of the United States. A recent study (Grugni et al. 2019) reports that the Q1b-Y4276 mutation is distributed from Siberia to South America. The same study suggests this mutation is a potentially useful marker for Algic languages.

Another large linguistic family of the Eastern Woodlands is Iroquoian. These languages are found in the vicinity of the Great Lakes and extend southwards along the Appalachian Mountains to Georgia. Bolnick et al. (2006) suggest, based on their analysis of the genetic data, that the putative homeland of Iroquois-speaking peoples is found in the southeastern United States. However, linguistic, and archaeological perspectives place the Iroquoian homeland in the Appalachian uplands, which encompasses a vast area from Pennsylvania to Georgia (Snow 2013: 358).

Snow (2013: 359-360) describes indigenous peoples that inhabited the Mississippi River valley and the lower Ohio River. These peoples include those that speak languages belonging to the Siouan-Catawban and Muskogean language families. According to Snow, around the year 1000 AD many of the Siouan-Catawban peoples, such as the Mandan, practiced what appears to be intensive agriculture. However, a drought around 1450 AD pushed some of the Siouan-Catawban people such as the Lakota onto the Great Plains where they abandoned farming altogether. This reversion from agriculture to foraging also occurred in Europe among the Uralic-speaking populations of the Baltic and Scandinavia (See Chapter 15).

5.4. The American Southwest.

The American Southwest cultural area consists of California, Nevada, Arizona, New Mexico, and southwestern Colorado. Linguistic diversity in this area includes the Uto-Aztecan, Kiowa-Tanoan, and Eyak-Athabaskan language families as well as the Zuni language isolate. Haplogroup Q-M242 data are available for the Jemez, Tohono O'odham (Papago), Akimel O'odham (Pima), Navajo, and Apache peoples (see **Supplementary Tables 17.3** and **Table 17.5**). C2b-P39 has also been reported for the Navajos and Apache (see **Supplementary Table 17.6**). The Jemez language belongs to the Kiowa-Tanoan language family. The Pima and Papago speak O'odham, an Uto-Aztecan language. Navajo and Apache are classified as Eyak-Athabaskan languages.

The Uto-Aztecan family has a vast geographical distribution, from Oregon in the United States to Panama (Campbell 1997: 133). This family consists of 61 languages. The two main divisions are Northern Uto-Aztecan and Southern Uto-Aztecan. Northern Uto-Aztecan consists of 13 languages found in the United States. Examples include Hopi, Comanche, Shoshoni, and Paiute. The Southern Uto-Aztecan branch consists of 48 languages. Forty-seven of these languages, such as Nahuatl, the language of the Aztecs, are found in Central America. O'odham, the language of the Pima and Papago, is the only Southern Uto-Aztecan language found in North America.

Agriculture may have played a role in the expansion of Uto-Aztecan languages. Additional details will follow below in the discussion of Central America. In the meantime, it is necessary to discuss Numic languages, a sub-branch of Northern Uto-Aztecan. Its speakers include the Comanche, Paiute, Mono, and Shoshoni peoples. According to the archaeological record, it appears as though they abandoned farming about a thousand year ago and adopted foraging as their subsistence strategy (LeBlanc 2013: 373).

5.5. Central America.

For the purposes of this discussion, the border of the United States and Mexico defines the northern boundary of Central America. The border of Panama and Colombia defines the southern boundary. In Central America, Y-chromosome population data are available for the following language families: Chibchan, Chocoan, Mayan, Mixe-Zoquean, Otomanguean, Tarascan, and Uto-Aztecan (see **Supplementary Tables 17.3**, **17.4**, and **17.5**.

All the Uto-Aztecan languages of Central America are classified within the Southern Uto-Aztecan branch. The Otomanguean family consists of 178 languages found in Mexico. Examples of Otomanguean languages include Mixtec, Zapotec and Otomi. The Mayan family consists of 31 languages found in Mexico and Guatemala. This language family is considered a linguistic relic of the Mayan civilization. Seventeen languages are classified within the Mixe-Zoquean language family of Mexico. Campbell and Kaufman (1976) suggest this language family is a linguistic relic of the Olmec civilization. The Chibchan family consists of 20 languages which are found in Costa Rica, Panama, Honduras, Nicaragua, and Columbia.

Bellwood (2005: 237-239) provides a short discussion of the co-evolution of farming and language in Central America. He suggests that early maize cultivation fueled expansion of the Mayan, Otomanguean, and Mixe-Zoquean language families. Bellwood (2005: 240-244) also discusses the Uto-Aztecan language family. He takes the position that the distribution of this language family follows an expansion of maize cultivation. This opinion was shaped by collaboration with the anthropologist Jane Hill. In a paper published in 2001 she suggests that Uto-Aztecan speakers were among the early maize farmers of Mexico. Around 6,000 years ago as the result of population pressure they began to expand northwards. Between 3,000 and 4,000 ago they migrated into the American Southwest and continued to cultivate maize and other crops. Hill supports her model mostly with linguistic reconstructions. A study from 2010 (Kemp et al) supports her model with Y-chromosome data, the distribution of Q1b-M3 and Q1b-Z780 mutations, and short tandem repeat (STR) data.

An alternative interpretation of the data analyzed by Kemp et al (2010) would posit the absence of a unique genetic signature for Uto-Aztecans or any of the other Central American language families. Available good resolution markers include Q1b-Y12421, which represents the majority of Q-M3 variation among Panamanians; Q1b-M924, which represents most of the Q-M3 variation in Mexico; Q1b-Z5906, which is distributed from Mexico to Argentina with a peak frequency in Peru; and Q1b-Z5908, which is distributed from Mexico to Argentina with a peak frequency in Peru (Grugni et al. 2019). At best, these recently reported markers merely suggest population growth in Central and South America population beginning about five thousand years ago. However, we cannot build farming-language expansion models with the currently available Y-chromosome data. Rather, the available data support *in situ* co-evolution of language and agriculture in Central America and not a co-expansion of language and agriculture.

The position taken by Hill (2001) is controversial. She places the origins of Proto-Uto-Aztecan in south-central Mexico where maize was first cultivated. Campbell, on the other hand, place the putative homeland of Uto-Aztecan languages somewhere in the southwestern United States or northern Mexico (1997: 150). Additionally, the Uto-Aztecan language-farming expansion, as posited by Hill (2001), was contested in a 2009 paper. Merrill et al. asserted that phonological reconstructions for flora and fauna place the putative homeland in Nevada and not in southern Mexico. Based on climatological data, the researchers further assert that a drought led to a bifurcation of Proto-Uto-Aztecan into the Northern and Southern Uto-Aztecan branches about 9,000 years ago. Southern Uto-Aztecan then expanded southwards from Nevada into Mexico. The researchers further suggest, based on their analysis of climatological and archeological data, that a Southern Uto-Aztecan group back-migrated from Mexico into the southwestern United States about 6,000 years ago. According to the report, this back-migration brought domesticated maize from Mexico into the region. Finally, Merrill et al. (2009) suggests that this expansion of maize and language was fueled by climate change rather than population pressure.

The cultivation of maize obviously stands as an important component of modeling the prehistory of Uto-Aztecan languages. This crop ultimately became an important food resource among many of the Native American cultures. It was the only grain-like food resource of the Western Hemisphere that can be stored for a long period of time. However, the road to a food staple was a long and complicated process that required considerable genetic modification of teosinte, the wild plant from which modern domesticated maize evolved. A study from 2018 (Kistler et al.) examined the domestication of maize using a synthesis of genetic, archaeological, and botanical data. The researchers suggest that domestication began roughly 9,000 years ago in south-central Mexico. However, according to the study, even 5,300 years ago the Mexican variant of maize had not evolved into a food staple. As such, the proposed timing of northward co-expansion of Uto-Aztecan and maize about 6,000 years ago, as suggested by Hill (2009), seems problematic because at this point in time maize could not have fueled reproductive success, which is an essential component of her hypothesis.

Figure 17.4. Teosinte (top), Teosinte-Maize Hybrid (middle), and Maize (bottom). Source: Wikipedia and John Doebley.

Hill's hypothesis is also undermined by evidence that suggests Maize was initially cultivated in the southwestern United States as a recreational crop rather than a food staple. Smalley and Blake (2003) provide a useful discussion of maize origins from botanical and anthropological perspectives. As previously mentioned, modern domesticated maize evolved from the wild teosinte plant. The report notes that teosinte cobs are much smaller than modern maize. Moreover, the kernels are barely edible. The study even describes teosinte kernels as "starvation" food that is otherwise "utterly useless." As such, this poses an interesting question: why would anyone waste so much time and energy to cultivate such a useless plant? According to Smalley and Blake (2003), the answer is alcohol. The teosinte stalks are sweet and initially people chewed them. Eventually someone discovered that when pressed the stalks yield syrup that can be used for corn wine. As such, people initially cultivated maize as a recreational product rather than for food. According to Smalley and Blake (2003), during the recreational phase of maize domestication, farmers planted seeds that they had gathered from the larger maize stalks with the idea of obtaining a larger yield of syrup with the next harvest. This selection of seeds from larger stalks eventually produced the large cobs that are characteristic of modern domesticated maize. At this point people started to dry maize kernels and maize became a food staple throughout the Americas. Additionally, some continued to produce alcohol from maize by using the kernels for making beer such as *chicha*.

5.6. Central Andes.

Heggarty and Beresford-Jones (2010) define the Central Andes region as the central Peruvian highlands and the western Pacific coastline of Peru. Researchers have proposed that the Quechuan and Aymaran language families evolved in this region (Bellwood 2005: 235; Heggarty and Beresford-Jones 2010). In terms of number of speakers, Quechuan represents the largest of Native American languages. According to *Ethnologue* (2019) around 7.8 million people speak one of the 44 Quechuan languages. Aymaran represents a smaller language family with three languages and around 1.7 million speakers. Although both language families share vocabulary and many structural similarities, mainstream linguistic opinion (e.g., Campbell 1997: 188) does not support the evolution of Quechuan and Aymaran from a common proto-language.

Very solid archeological and genetic evidence place *Homo sapiens* in South America by at least 10 thousand years ago (Roosevelt et al. 1996; Moreno-Mayar et al. 2018; Capriles 2019). From a Y-chromosome perspective the genetic relics of this migration are the Q1b-M3 and Q1b-Z780 mutations (see **Supplementary Tables 17.3** and **17.4**). Downstream from Q1b-M3 marker, several mutations point to substantial population growth in the Central Andes within the last five thousand years (Jota et al. 2016; Grugni et al. 2019). Agriculture probably fueled the expansion.

In order to understand the agricultural transition in the Central Andes, a discussion of geography is necessary. One finds a very steep rise in elevation. The western coast of Peru lies at sea level. Further east, in the central highlands, the elevation can reach 6,000 meters. During the Pre-Ceramic phase, about 11,000 to about 4,000 years ago, human activity was concentrated along the coastline (Heggarty and Beresford-Jones 2010). Abundant marine resources appear to have drawn people to this area. Nevertheless, archeological remains, radio-carbon dating, and stable oxygen isotope data (Haas et al. 2017) suggest that coastal hunter-gathers made seasonal treks into the highlands to harvest wild camelids such as alpacas. Then around 7,000 years ago, humans occupied the highlands on a permanent basis.

Figure 17.5. Lamas. Source: Wikipedia and Kallerna

It is important to emphasize that the agricultural transition in the Central Andes was a gradual process that involved increased dependence on domesticated crops and animals, and less dependence on foraging, a process that lasted several thousand years. The domesticated animals include alpacas, vicuñas, alpacas and llamas, camelids that were once hunted and as noted above, the incentive that initially drew people into the region. They provided a reliable source of meat as well as fleece for clothing. Additionally, highlanders utilized these animals as beasts of burden (for more details, see Mengoni-Gonalons and Yaco-Baccio 2006).

The domesticated plants include potatoes, sweet potatoes, quinoa, and maize. Since modern potatoes consists of many hybrids and variants, identifying how and when this food resource arrived in the Central Andes is difficult. The genetic data (Spooner et al. 2005; Hardigan et al. 2017) suggest that this tuber was initially domesticated in southern Peru about 8,000 to 10 thousand years ago. The sweet potato also became another tuber that was cultivated in the highlands. Like the potato, it is difficult to determine how and when the sweet potato arrived in this region due to the numerous hybrids and variants. Researchers suggest that this crop evolved independently in the Caribbean, Central America, and northwestern South America (see Roullier et al. 2013).

Quinoa, a type of chenopod, also became an important crop resource in the highlands. This plant is sometimes confused as a grain product, but it is technically a source of edible seeds. A study suggests that this crop was domesticated in the region about 3,000 years (see Bruno 2006). Finally, maize became an important food resource in the Central Andes around 3,000 years ago. This crop was cultivated both in the highlands and lowlands. Heggarty and Beresford-Jones (2010) suggest that the cultivation of this crop signals an intensification of agriculture within the regions. Bellwood (2005: 163-164) suggests that this intensification of agriculture eventually produced population pressure and soil degradation along the coast. As a result, coastal settlements were abandoned, and the highlands became the focal point of human activity.

Figure 17.6. Peru and the Andes Mountains.

Returning to the linguistic prehistory of the Central Andres, language diversity in the region was shaped by several different cultural transitional periods that arose between the adoption of maize (around 3,000 years ago) and the arrival of the Spanish in 1532. Heggarty and Beresford-Jones (2013: 405) describe Aymaran languages as a linguistic relic of the Chavin culture and the Early Horizon period, roughly 900 BC to 100 AD. Both researchers describe Quechuan as a linguistic relic of the Wari civilization and the Middle Horizon Period, roughly 550 AD to 1000 AD.

In their 2010 paper, Heggarty and Beresford-Jones consider Bellwood's early farming dispersal hypothesis (2005: 1-11). They acknowledge that this model helps to decipher linguistic evolution in the "Old World." However, according to the researchers the model is problematic in the Central Andes. Instead of a co-expansion of language and early agriculture, as predicted by the early farming dispersal model, linguistic variation in the Central Andes conforms to a model of *in situ* co-evolution of language and farming.

The *in situ* co-evolution of language and farming is a prehistoric language trajectory that also occurred in the Old World. Japonic and Koreanic are two examples (see Chapter 15). However, the transition to intensive agriculture leveled linguistic diversity in both regions. In the Central Andes, on the other hand, the same leveling of linguistic diversity seems not to have occurred. Rather, Heggarty and Beresford-Jones (2010) suggest diglossia within the region. When the Spanish arrived, Quechua was the high variety and Aymara the low. Perhaps this diglossia reflects the absence of intensive agriculture for a sufficient period of time. Extending this argument further, the Spanish may have interrupted what would have ultimately been a natural leveling of language diversity that is characteristic of intensive agriculture.

The idea that agriculture follows a gradient of intensification was explored by Stevens and Fuller in their 2017 paper. They suggest that the transition to agriculture only occurs when a population obtains 50 percent of its calories from domesticated plants and animals. According to the report, the road to agriculture can have a lengthy pre-agricultural phase. During this phase, hunter-gathers often cultivate crops on a smaller scale. However, this is not agriculture. Rather, as suggested by the study, the transition to agriculture essentially marks a point-of-no-return. Agriculture vastly improves reproductive success, and this comes with a price. At this point foraging is no longer an option because you must feed many more people. Furthermore, habitat for wild animals and plants are now utilized as farmland. From a linguistic perspective, it seems that intensive agricultural eventually creates social institutions that undermine linguistic diversity.

5.7. Amazonia and the Caribbean.

Amazonia is usually associated with the world's largest rainforest. For the purpose of this discussion, the geography of this region is defined by the Orinoco and Amazon Rivers and the vast number of tributary rivers that flow into them (see Figure 17.7 below). The archaeological record suggests that Amazonia has been inhabited for at least 10 thousand years (e.g., Roosevelt et al. 1996; Capriles 2019). This closely follows dating estimates acquired from ancient DNA data retrieved from the Sumidouro Cavern in Brazil (Moreno-Mayar et al. 2018), which provide the most robust time estimates for the human settlement of South America.

Amazonia is linguistically complex. Major language families of the region include the Carib, Tupi, Panoan, Jean, Tucanoan, and Arawak language families. This discussion will focus on Arawak. Interestingly, the prehistory of this language family is strikingly similar to that of Austronesian. Moreover, Arawak provides an especially good example of a New World language family that conforms to the *early farming dispersal hypothesis* as postulated by Bellwood (2005: 1-11).

The Arawak language family is also known as Maipurean. *Ethnologue* (2019) lists 56 different Maipurean languages spoken by roughly three quarters of a million people. Maipurean has two main divisions, a northern branch and a southern branch. Southern languages are found in Peru, Bolivia, and Brazil. The northern branch is found in Brazil, Suriname, Guyana, Columbia, Venezuela, Puerto Rico, and Honduras. Prior to 1492 Arawak was a linguistic heavyweight within Amazonia and the Caribbean. Arawak languages thrived and survived because the Arawak people had mastered the art of tropical agriculture along major river systems. Part of their success stems from the construction of raised field agriculture. Amazonian rivers tend to flood regularly. By constructing fields above the floodplain, they greatly increased the efficiency of agriculture by ensuring adequate drainage and improving the fertility of otherwise poor growing soil (Whitney et al. 2014). Another factor is crop selection which includes sweet potatoes and maize. However, the most important crop was manioc, which like the potato and sweet potato, is also a tuber. Also known as cassava, manioc can be cultivated in the poor-quality soil of tropical climates. Another advantage is that bitter manioc can be made into a flour and stored.

The archeological and botanical data (Whitney et al. (2014) place the origins of raised field agriculture in the Llanos de Moxos region of Bolivia about 2,500 years ago. This dating estimate for raised field agriculture corresponds to a massive expansion of Arawak settlements alongside the numerous rivers of Amazonia (Horborg 2005; Heckenberger 2013). Furthermore, the South American variant of domesticated maize traces its origins within or near this region (Kistler et al. 2018). Finally, the region is a potential domestication center for manioc (Olsen and Schaal 1999).

Figure 17.7. Amazon River Basin. Source: Wikipedia and Kmusser.

Similar to the "Austronesian advantage" that evolved in Island Southeast Asia, the "Arawak advantage" evolved in South America. The so-called "Arawak Advantage" points to cultural adaptations, such as raised field agriculture and plant domestication, that drove greater reproductive success in the tropical rainforest. This success then drove an expansion of Arawakan languages. Interestingly, some researchers have asserted that foraging cannot sustain human population in tropical rainforests such as Amazonia. They maintain that to survive in these regions, people need to supplement their diet with agriculture (Bailey et al. 1989). However, this position seems contrary to the archaeological data. Roosevelt (1996), for example, provides evidence of pre-agriculture occupation of Amazonia. Nevertheless, one still finds compelling arguments for the idea that foraging is not capable of sustaining high population density within tropical rainforests. Thus, it seems significant in that Horborg (2005) describes the Pre-Columbian Arawak settlements or villages as chiefdoms with high population density.

When Columbus landed in the New World, among the first indigenous peoples he encountered were the Taínos, speakers of the Taíno language, which belongs to the Arawakan language family. Taínos were descendants of a second Arawak expansion that began roughly 2,500 ago from the northern coast of South America. Initially, the Taínos settled the Lesser Antilles Islands. They later expanded into Hispaniola, Puerto Rico, the Bahamas, Jamaica, and Cuba. This expansion carried many of cultural features of Arawak cultures found on the South American mainland: the intensive cultivation of manioc; the dominance of large trade networks; villages centered around a plaza; pottery; social organization; and high population density (see Wilson 2007: 59-136; Keegan 2013: 376-383 for a more detailed discussion).

Figure 17.8. The Caribbean.

Dixon and Aikenvald (1999: 7) estimate that between 2 and 5 million people lived in Amazonia prior to the arrival of Europeans. According to the report, since 1492 European diseases and population displacement have significantly altered the cultural and linguistic landscape of this region. Consequently, it is difficult to reconstruct the prehistoric linguistic diversity of the region. The researchers further note that the surviving indigenous languages of the Amazonia remain understudied in academia. Similar to the paucity of linguistic data, one finds a very limited amount of Y-chromosome data. The published data mostly consists of frequency results for the Q1b-M3 and Q1b-Z780 mutations, low resolution markers that are not particularly informative (see **Supplementary Tables 17.3** and **17.4**). Furthermore, Native American Y-chromosome lineages have disappeared among contemporary Caribbean populations (e.g., Marcheco-Teruel et al. 2014).

A recent whole-genome study (Fernandes et al. 2021) presented data taken from human remains from the Bahamas, Haiti, the Dominican Republic, Puerto Rico, Curaçao, and Venezuela. Most of the Y-chromosome data belongs to the Q1b-M3. However, Q1b-Z780 was found in two remains from the Bahamas (see **Supplementary Table 17.6** for more details). Important conclusions from the study include evidence of admixture between the Archaic culture that expanded into the Caribbean about 6,000 years ago and the Arawaks who arrived about 2,500 years ago. Additionally, the data support exchange networks between the Caribbean islands. Finally, the data eliminate North America as source of Arawak expansion, but South and Central America are equally possible.

As previously noted, the prehistory of Austronesian and Arawakan languages is strikingly similar. Austronesians and Arawaks excelled at navigation. As a result, both groups dominated regional trade alliances. Austronesians and Arawaks excelled at tropical agriculture. Both groups cultivated tubers that grow in the tropic environment: taro in the case of Austronesians, and manioc in the case of

Arawaks. Both groups farmed were nobody could farm. An evolutionary adaptation allowed Austronesians to farm in malaria infested lowland coastal regions. Arawaks perfected riverine agriculture by constructing raised fields above the flood plain. For Austronesians and Arawaks, agriculture became a successful cultural adaptation that fueled rapid population growth, which fueled a rapid co-expansion of people and language.

Section 6. Athabaskans.

The reader should note that the terms Na-Dené and Eyak-Athabaskan are essentially synonymous. Greenberg (1987: 321-330), for example" describes the so-called "Na-Dené problem." However, other linguists such as Campbell (1997: 110-155) and *Ethnologue* (2019) use Eyak-Athabaskan. According to Ethnologue (2019), the Eyak-Athabaskan language family consists of 44 different languages. These languages have a leap-frog distribution over a vast geographic range (see Figure 17.8 below).

Ethnologue (2019) divides the Eyak-Athabaskan language family into three main divisions: Eyak, Athabaskan and Tlingit. Eyak is a single language that is now extinct. It evolved near the mouth of the Copper River in southern Alaska. Tlingit is a single language branch from the coastal region of southeastern Alaska. The Athabaskan branch has three sub-branches: Apachean, Northern Eyak-Athabaskan, and Pacific Coast Eyak-Athabaskan. Apachean languages are found in the desert of the southwestern United States. This sub-branch consists of the Navajo and Apache. The Northern sub-branch consists of 27 different languages distributed throughout Alaska and Canada. Finally, the Pacific Coast sub-branch consists of languages along the Oregon and Californian coast in the United States.

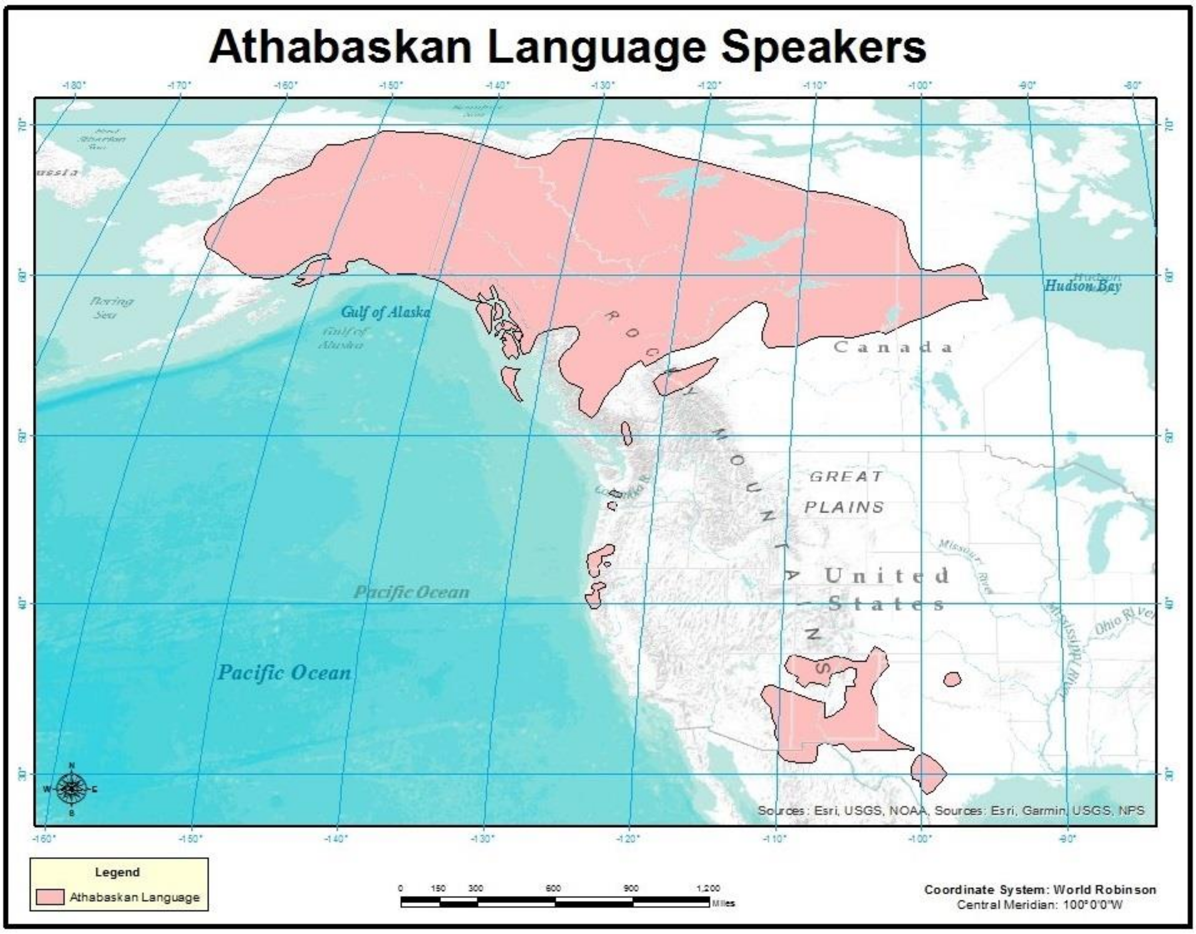
From a linguistic perspective (Campbell 1997: 110-115), the Eyak-Athabaskan language family potentially arose within the interior of North American, where Alaska, British Columbia and the Yukon converge on the map. Eskimo influence in Eyak-Athabaskan is negligible, however, despite close geographic proximity in Alaska and Canada. The researcher also suggests that Haida and Eyak-Athabaskan borrowed from each other. However, the data fail to support a genealogical relationship for both language families.

From an archeological perspective, Gillispie (2018) suggests that the Athabaskan cultural tradition arose in interior Alaska about 1,700 years ago. According to the researcher, the appearance of the cultural tradition corresponds to a technological innovation in the region, the bow and arrow. Interestingly, Gillispie (2018) suggests Eyak, Tlingit, and Haida societies evolved before the Athabaskans, at around 2,500 years ago. This estimate corresponds to climate change that stabilized coastlines, as well as cooler weather and greater precipitation. Thus, Eyak-Athabaskan potentially evolved along the southern Pacific coast of Alaska rather than in the Alaskan and Canadian interior.

Matson and Magne (2013) time the Athabaskan expansion into interior Alaska and British Columbia with the Mount Churchill volcano eruption around 300 AD. Both researchers suggest that a second more powerful eruption occurred around 800 AD. This eruption drove Athabaskan peoples either into the northwestern Pacific coast of the United States or into the American Southwest. Linguistic support for this expansion model comes from the Dakelh people of British Columbia. They speak Carrier, an Athabaskan language that is closely related to Apachean languages.

An interesting report from 2021 (Kristensen, Ives, and Supernant) presents paleoenvironmental, historical, and archeological data from the Mount Churchill explosion that occurred 1,200 years ago. According to the researchers, when the explosion occurred, the Athabaskans had mediated trade between coastal peoples and indigenous peoples in the interior of North America. Among the trade commodities was copper, a valuable resource for making ornaments, knives, and points. Through trade networks, the Athabaskans eventually established extensive kinship networks in the northwestern part of North America. When the explosion scattered a large volume of volcanic ash over parts of Alaska and the Yukon and Northwest Territories, the Athabaskans fled and sought refuge among trading partners and kin who lived beyond the devastation.

Figure 17.9. Distribution of Athabaskan Languages. Source: National Park Service.



Very little contemporary Y-chromosome data exists for Eyak-Athabaskan populations see **Supplementary Tables 17.3**, **17.4**, and **17.5**). A 2012 study (Schurr et al.) considered the long-standing linguistic debate about the classification of Haida and Eyak-Athabaskan. The study was not able to provide genetic evidence of a common ancestral population for these populations. Nevertheless, this study along with others, report the presence of the rare C2b-P39 mutations among Eyak-Athabaskan populations which include the Tlingit of southeastern Alaskan coast; the Tanana of interior Alaska; the Dogrib and Gwich'in of Canada; and the Navajo and Apache of the southwestern United States.

It should be emphasized that C2b-P39 has also been detected in Algic, Eskimo-Aleut, Iroquoian, Muskogean, and Siouian-Catawban speaking populations (see **Supplementary Table 17.5** for additional details). Given the contemporary distribution of C2b-P39, this mutation represents a potential founder lineage for North America. In other words, Upper Paleolithic hunter-gatherers may have carried the mutation when they crossed over the Bering land bridge roughly 15 thousand years ago. Such a position was taken by Wei et al. in their 2018a report that analyzed contemporary Asian C2-M217 lineages closely related to C2b-P39. Ancient DNA from Brazil also support this position. The C2b-L1373 mutation was sequenced from 10-thousand-year-old remains from Lapa do Santo in Brazil (see Chapter 7 for more details). However, an alternate scenario would suggest that C2b-P39 represents more recent geneflow between Alaska and the Kamchatka Peninsula as suggested by Pinotti et al. (2019). Such a scenario is supported by the presence of the C2b-FGC28881.2 mutation found in Koryaks (Wei et al. 2017). The C2b-FGC28881.2 mutation is the closely related to the C2b-P39 sister clade mutation.

C2b-P39 data from contemporary populations, along with Q-M242 data from contemporary and ancient DNA studies, potentially support prehistoric contact between the indigenous peoples of northeastern Asia and the Athabaskans of Alaska. Prehistoric contact is a salient point for linguists because some researchers have proposed a common proto-language that unites the Yeniseian language family of south-central Siberia and the Eyak-Athabaskan language family of North America. The so-called "Dene-Yeniseian connection" was initially proposed by Merritt Ruhlen in 1998 based on 36 cognate sets which include basic vocabulary. This hypothesis is controversial among the linguists. Campbell (2011) asserts a lack of linguistic evidence. The geographic distance between Yeniseian and Athabaskan speaking populations is also a problem.

A whole genome (autosomal) study (Flegontov et al. 2016b) endorsed the so-called Dene–Yeniseian hypothesis. The model proposes a common proto-language that unites the Eyak-Athabaskan and Yeniseian language families. However, this conclusion is inconsistent with the Y-chromosome data. It should be noted that the Ket tribe of south-central Siberia are the sole source of genetic data for Yeniseian languages. Turning now to the Y-chromosome perspective, as illustrated by **Supplementary Figure 17.1**, the Q1b-L54 mutation splits into Q1b-M930, Q1b-Z780 and Q1b-L330. The Q1b-L330 mutation is not found in Native Americans. Rather, it is confined to Siberia where it is the predominate mutation of the Ket people (e.g., Flegontov et al. 2016b). Dating estimates provided by Wei et al. (2018) suggest that if Kets and Native Alaskans evolved from a common ancestral population, then both groups separated more than 17 thousand years ago. Given the great time depth, 17 thousand years, linguistic tools cannot prove a common proto-language Eyak-Athabaskan and Yeniseian.

Finally, the linguist Joseph Greenberg suggests (1987: 323) that Amerind represent an initial migration into the Americas, and that the Eyak-Athabaskan represents a second migration much later. Based on previously detailed archaeological and linguistic perspectives, Greenberg's proposal seems unlikely. Nevertheless, anthropological, and genetic perspectives (see Sections 7, 8, and 9 below) present evidence of recent cultural contact between Native Alaskans and the indigenous peoples of Eastern Siberia. As such, language contact could potentially explain similarities found in Yeniseian and Eyak-Athabaskan languages. However, this requires evidence that places Yeniseian-speaking people along the coastline of eastern Siberia, which is clearly not the case.

Section 7. Eskimo-Aleut.

According to *Ethnologue* (2019) the Eskimo-Aleut language family consists of 11 languages. This language family has two main divisions, Aleut, a single language branch, and the Eskimo branch with the ten remaining languages. The Aleut language is found on the Aleutian Islands. Eskimo has two sub-branches, Inuit-Inupiaq with five languages and Yupik with five languages. The geographic distribution of Inuit-Inupiaq languages follows the Alaskan coastline north of Unalakleet along the Bering Sea and Artic Ocean. They further extend along the Arctic Ocean coastline into Canada and Hudson Bay. From Hudson Bay, Inuit-Inupiaq extends into Greenland. The Yupik sub-branch is found on both sides of Bering Sea. Two Yupik languages, Naukan and Sirenik, are spoken on the Chukotka Peninsula in Russia. Three Yupik languages are spoken the United States: Central Yupik language along the Bering Sea Coastline of western Alaska; Pacific Yupik along the Pacific coastline of southern Alaska; and St. Lawrence Yupik on St. Lawrence Island in the Bering Sea.

For those interested in more details about Native Alaskan languages, the Alaska Native Language Center at the University of Alaska is a tremendous resource. Their website can be accessed with this link: **https://www.uaf.edu/anlc/**. Additionally, their researchers have prepared a detailed map that depicts the distribution of Athabaskan and Eskimo Aleut languages, which can be accessed with this link: **Map Native Alaskan Languages**.

In his reference guide to Native American languages, Campbell (1997: 109), places the geographic point of origin of Eskimo-Aleut languages in southwestern Alaska near Bristol Bay and the Cook Inlet. In the discussion he rejects a close linguistic relationship between Eskimo-Aleut and the Uralic family of Northern Eurasia. Similarly, he finds a common ancestral language for Eskimo-Aleut and Chukotko-Kamchatkan languages problematic.

Turning now to the archaeological perspective, the where and when of Eskimo-Aleut languages origins is confusing. Does the archeological record link this language family with Paleo-Eskimos or Neo-Eskimos? The Paleo-Eskimos are often identified with the Dorset cultural tradition. About 5,000 years ago, the Arctic small tool tradition evolved in circumpolar region of northeastern Russia (Tremayne 2015). A secondary expansion of the Arctic small tool tradition, the Dorset culture, then expanded across the Arctic Ocean of North America, from Alaska to Greenland. This occurred by around 3,000 years ago.

As previously stated in Chapter 15, about 50 percent of Siberian Yupik have the N-M231 haplogroup whereas the mutation is absent among North American Eskimos. Perhaps the arrival of reindeer herders in northeastern Siberia and population pressure drove some of the Paleo-Eskimos across the Bering Sea into Alaska and beyond. This would assume that Eskimo-Aleut peoples predate the arrival of reindeer domestication in northeastern Siberia. After crossing the Bering Sea, the archaeological record (Gillispie 2018: 30) suggests that the Paleo-Eskimos of North America were highly mobile foragers. They alternated their subsistence strategy between inland and coastal resources. In the winter they settled along the coast to hunt seals. When the weather became warmer, they moved inland to intercept migrating herds of caribou and muskoxen.

The Neo-Eskimo or Thule cultural tradition eventually replaced the Paleo-Eskimo tradition. Fortescue (2013: 341) and Gillispie (2018: 23) suggest that this occurred by around one thousand years ago. This transition involved significant cultural changes. The Thule became successful whale hunters. As a result of this food resource, they built permanent settlements and focused on marine resources. This resource also helped to increase population density. Permanent settlements and greater population density eventually brought more complex social structures, and with that, trade alliances and warfare (Friesen 2013: 349-351).

According to Fortescue (2013: 340) linguistic relics of the Paleo-Eskimos have disappeared. This suggests that the Thule tradition involved a population expansion and potential assimilation of the Paleo-Eskimos. Eskimo-Aleut peoples and languages then radiated in several directions: westward into the Aleutian Islands and northeastern Asia; eastward along the southern Alaska coast; northwards along the eastern Bering Sea coastline of Alaska; and finally, along the Arctic Sea coastline of North America. However, the available genetic data detailed in Section 8 (below) paint a different picture. Eskimo-Aleut languages may extend much further back in time to the Paleo-Eskimos.

Section 8. The Bering Sea and Exchange of Language, Genes, and Culture.

8.1. The Linguistic Data.

As discussed previously in Section 6, some researchers suggest that a common ancestral language may link Eyak-Athabaskan and Yeniseian. However, this proposal is also controversial among the linguists. Far less controversial is the idea that the Eskimo-Aleut family was shaped by the indigenous peoples of North America and northeastern Asia. Linguists agree that Eskimo-Aleut languages are spoken on both sides of the Bering Sea. Moreover, linguistic evidence may suggest that Eskimo-Aleut may have been spoken on the Kamchatka Peninsula by the coastal Chukchi people (Fortescue 2004). It should be noted that they now speak Chukotka-Kamchatkan languages.

8.2. The Anthropological Perspective.

From an anthropological perspective, whale hunting appears to have mediated long-term cultural exchange along the northern Pacific Rim. Savelle and Kishigami (2013), in their discussion of prehistoric subsistence whaling, draw a distinction between opportunistic and active whaling. According to the researchers, Jomon archeological sites in Japan provide evidence of opportunistic whaling at around nine thousand years ago. Intensive or active whaling then began about 5,000 years ago on the Noto Peninsula in Japan. Active whaling eventually spread northwards through the Kurile Islands. By around 3,000 years ago, active whaling reached the Kamchatka Peninsula and Chukotka (see Figure 17.4). Perhaps as early as 1,500 years ago, whaling reached Alaska. Finally, by around 800 years ago, whaling had advanced across northern Canada.

Heizer (1944) presents a report that explores whaling methods across the northern Pacific Rim. The Jomon people of Japan used nets to harvest whales. However, the Ainu of the Kurile Islands and southern Kamchatka Peninsula utilized a dart or lance that had been coated with aconite poison. An individual or small hunting party paddled out to sea and stabbed a whale just one time. The poison eventually killed the whale. Hunters then waited for the dead animal to float ashore. This Ainu method of harvesting whales was later adopted by the Aleutian Islanders and the Alutiiq (Pacific Yupik) on Kodiak Islands. This suggests that the Aleutian Islands facilitated linguistic, cultural, and genetic exchange between Alaska and northeastern Asia.

In his 1944 report, Heizer stated that the Koryak of the Kamchatka Peninsula employed a much different method of harvesting whales compared to that employed by the Ainu. The Koryak method utilized a larger hunting party and large boats. The hunters rowed out to sea and stabbed a whale repeatedly with harpoons. The harpoon had a detachable point that affixed a line and a float to the whale. Eventually the whale succumbed to wounds and exhaustion. Then it was towed ashore. The Koryaks method later spread to the Chukchi and the Asian Eskimos (Yupik) of Chukotka, and then across the Bering Sea, where it was adopted by Alaskan Eskimos along the Arctic Ocean.

Figure 17.10. Japan, Russia, and Alaska.
Source: University of Texas.

8.3. The Genetic Evidence.

Researchers have identified three high resolution mutations that may support the bidirectional spread of culture, language, and genes across the Bering Sea within the last 5,000 years: Q1a-B143, Q1a-B277, and Q1b-B34. It should be noted that the Q1a-B143 mutation is a downstream variant of the Q1a-F746 mutation. The Q1a-F746 mutation was identified in remains found at the Afontova-Gora-2 archaeological site along the Yenisei River in the Altai-Sayan region (see, also, **Supplementary Table 17.6**). These remains are from a man who died roughly 17 thousand years ago.

The Q1a-B143 mutation is observed among contemporary populations in Alaska and Siberia (Karmin et al. 2015; Grugni et al. 2019). Among ancient populations, the oldest Q1a-B143 sample was recovered at the Duvanni Yar archaeological site in Siberia from the Kolyma-1 man, who died about ten thousand years ago. In the Americas, the oldest Q1a-B143 sample comes from Greenland and the so-called "Saqqaq man," a Paleo-Eskimo who died about four thousand years ago. These data from Siberia and Greenland support the linguistic and anthropological perspectives that place the origins of the Paleo-Eskimos among the Arctic small tool tradition of northeastern Eurasia. Additional support for this position comes from dating estimates provided by Sun et al. (2019). Although the researchers did not date Q1a-B143, they provide an estimate for Q1a-M120, a phylogenetically equivalent mutation. The estimate for Q1a-M120, about 7,000 years, strongly links the presence of Q1a-B143 in Americas with more recent geneflow within the last 5,000 years rather than the initial human expansion over the Bering land bridge.

The Saqqaq data also underscores the idea that Amerindians are genetically distant from Eskimo-Aleuts. Eskimo lineages include those that are downstream from Q1a-F1096 as well as Q1b-M346. Amerindian lineages, on the other hand, are downstream from Q1b-M346 and do not include Q1a-F1096 lineages (see **Supplementary Figure 17.1**).

Turning now to the Q1a-B277 mutation, data for this marker come from Eskimo and Athabaskan archaeological sites in Alaska, and Yupik remains from Siberia (see **Supplementary Table 17.6**). As reflected by **Supplementary Figure 17.1**, Q1a-B277 is downstream from Q1a-M25. Asian origins of Q1a-B277 are supported by the contemporary distribution of the Q1a-M25 mutation (see **Supplementary Table 17.7**) which includes Turkmen in Afghanistan and Mongols in western Mongolia. The oldest Q1a-B277 sample comes from Ust'-Belaya man, a Yupik who died 4,200 years ago in Chukotka, Russia. The presence of this mutation among Eskimo remains in Alaska provides additional support for the Arctic small tool tradition as the source of Eskimo-Aleut languages. Additionally, the Q1a-B277 data among Athabaskan remains suggest prehistoric gene flow between this population and Alaskan Eskimos.

The Q1a-B143 and Q1a-B277 mutations reflect geneflow from northeastern Asia into Alaska that began at least four thousand years ago. However, geneflow across the Bering Sea was not unidirectional. Rather Q1b-B34 data suggest that geneflow also occurred in the opposite direction, from Alaska to northeastern Asia. In order to better understand the evolutionary history of this mutation, the reader is directed to **Supplementary Figure 17.1**. As shown by the figure, Q1b-M3 splits into Q1b-M848 and Q1b-Y4276. Downstream from Q1b-Y4276 is Q1b-B34. Grugni et al. (2019) outline several salient points about the Q1b-Y4276 mutation. First, it is distributed from Siberia to South America. Second, the mutation evolved in the Americas about 9,300 years ago. Third, the Q1b-B34 downstream mutation represents a back migration of Native Americans into northeastern Asia. Fourthly, based on dating estimates obtained from Koryaks, the back migration occurred about five thousand years ago. Finally, ancient DNA supports the back migration (see, also, **Supplementary Table 17.6**).

8.4. Conclusions.

Caution against drawing conclusions from a limited dataset is recommended. More data for Native Alaskans and indigenous peoples of Siberia are needed. It should be noted that resequencing of published Q1a-F746 data for Native Alaskan may also yield additional contemporary Q1a-B143 data (see **Supplementary Table 17.8**). Resequencing of Q1-F903 data for the Yupik, Chukchi, and Koryaks may provide additional Q1a-B143, Q1a-B277, and Q1b-B34 data for northeastern Siberia (see **Supplementary Table 17.9**). Additional samples from contemporary Native Alaskans and the indigenous peoples of northeast Asia could also clarify cultural exchange between both regions. The currently available data strongly support cultural exchange between Alaskan Eskimos and indigenous northeast Asians during the last five thousand years. Additional data may also yield evidence of contact between northeast Asians and Athabaskans.

Section 9. Problematic Models of Native American Origins.

9.1. Overview.

This chapter defines the term "Native American" as a cultural tradition that began roughly 11 thousand years in the Americas following the demise of the mammoths. Nevertheless, archaeological, and genetic debate has traditionally focused on the origins of Native Americas. Much of the research arising from this debate is useful for linguists, especially those interested in a triangulated Y-chromosome perspective of Native American languages. However, archaeological and genetic debate has also generated problematic models of Native American origins. As such, the goal of this discussion to warn linguists of several pitfalls that have surfaced.

9.2. The Genetic Ancestry of Native Americans.

The term "whole genome" reflects attempts to use autosomal data as a tool for deciphering human genetic history. As detailed previously in the first chapter, autosomal research utilizes alleles rather than mutations as a genetic tool, whereas mtDNA and Y-chromosome data utilize mutations that are found on non-recombinant regions of the human genome. As such, analysis of the autosomal data requires complex statistical analysis to overcome the reshuffling of genetic cards that occurs as the result of recombination. mtDNA and Y-chromosome data, on the other hand, overcome this problem as they gathered from non-recombining regions of the genome. Y-chromosome data become the tool of choice because the larger size of this locus provides a much more detailed picture of genetic variation.

Genomic reports have surfaced that posit ancient Asians as the genetic ancestors of contemporary Native Americans. Raghavan et al. (2014), for example, report data for the so-called Mal'ta boy. He was a two-year-old child who died along the banks of Belaya River in southern Siberia about 24 thousand years ago. The study suggests that Native Americans derive a significant part of their genetic ancestry from this individual. Another example is the report from Sikora et al. (2018). They present data for Kolyma-1, an individual who died ten thousand years ago in northeastern Siberia. The study suggests that Native Americans derive part of their genetic ancestry from this man.

The genetic history of Native Americans is a legitimate research question, and their genetic history certainly has a North Eurasian component. However, it should be emphasized that genes do not define ethnicity. Rather, the question of identity is a matter that Native Americans should define for themselves. Thus, it is important for researchers to differentiate genetic history from ethnicity. Mal'ta and Kolyma-1 might be genetic ancestors, but they are not cultural ancestors. Moreover, researchers should also consider whether the data support the position that Mal'ta and Kolyma-1 are, in fact, genetic ancestors of Native Americans. This posits an over-expansive interpretation of the data. A more conservative treatment of the Mal'ta data suggests that he merely represents part of the genetic inventory (or genome) of those living in southern Siberia at the time of the Last Glacial Maximum. Kolyma-1, on the other hand, merely represents part of the genome of northeastern Siberia at the beginning of the Holocene.

The geneticist Pavel Flegontov (2016a, 2016b, 2017, and 2019 has co-published several papers that explore the genetic history of Native Alaskans and the indigenous peoples of Siberia from a whole genome perspective. The 2019 paper presents, from an autosomal perspective, a model of bi-directional gene flow similar to the Y-chromosome model presented above in Section 8. Nevertheless, these papers illustrate the limitations of genomic tools for modeling the prehistory of language. The paleo-genomic models presented by these studies utilize complex statistical tools for deciphering question of "ancestry." Extending the argument presented above, this is a cultural matter that Eskimos, Athabaskans, Koryaks, Kets, and Yupik should decide for themselves. Triangulated Y-chromosome-based models, and the refreshing absence of statical analysis, clearly provide a more transparent methodology for deciphering the language prehistory of Native Americans and indigenous Siberians. Moreover, these models conveniently sidestep the thorny question of ethnic identity.

9.3. Beringian Standstill.

Those that explore the archeology and genetic history of Native Americans will certainly encounter the term "Beringian standstill." This model was initially proposed by Tamm et al. in their 2007 report that investigated the speed of the initial human migration wave into the Americas. They considered whether it was a rapid "direct colonization" event or, alternatively, if humans congregated in a refugium near the Bering land bridge before migrating into the Americas. Based on a comparison of Asian and New World mitochondrial DNA lineages, the researchers favored an "incubation" period, meaning that the first humans in America underwent genetic isolation for up fifteen thousand in a northeastern Asian refugium before migrating over the land bridge into Alaska.

The Beringian standstill hypothesis cannot be defended with archeological data (e.g., Buvit and Terry 2016; Potter et al. 2018). The hypothesis is also problematic from a Y-chromosome perspective. As suggested by Wei et al. 2018a, Y-chromosome diversity downstream from Q1b-L54 posits a rapid human migration from south-central Siberia into Alaka beginning about 16 thousand years ago.

9.4. Pre-Clovis Human Migrations into the Americas.

Section 3 of this present chapter provides a robust settlement model of the Americas that is well supported by a synthesis of archeological, climatological, genetic and Y-chromosome data. In summary, *Homo sapiens* crossed over the Bering Sea into Alaska roughly fifteen thousand years ago. By around 14 thousand years ago, they had migrated southwards through an Ice-Free Corridor onto the Great Plains of North America. Around ten thousand years ago, they arrived in South America.

As previously discussed in this chapter, archaeologists have traditionally used the term "Clovis" to define the first Native American culture. This term stems from a unique type of spear point that appeared roughly 13 thousand years ago (see Section 4). Archeological studies periodically surface that present evidence of "pre-Clovis" migrations. Bourgeon, Burke and Higham (2017) report, for example, human presence in North America by around 24 thousand years ago. This is based on cut marks on bones found at the Bluefish Caves site in the Yukon of Canada. According to the researchers, the cut marks were clearly made with human-made tools. Yet the study does not account for the possibility that scavengers made the cut marks ten thousand years after the animals had died. Conditions are such in the Arctic that animal remains are well preserved in ice for thousands of years.

Dillehay et al. (2015) presents another example of pre-Clovis migrations. Based on artifacts found at the Monte Verde archeological site, they suggest that humans arrived in southern Chile around fourteen thousand years ago. However, the use of artifacts to date human presence in an area can be problematic. Organic matter around the site provides the material for carbon-14 dating, and sometimes the results provide a poor correlation with human occupation. Dating results from human remains provide the most robust estimates for human occupation. Thus, Monte Verde dating estimates might be problematic because human remains have not been found at this location.

A recent study (Bennett et al. 2021), one that gained a fair amount of media attention, maintains that *Homo sapiens* were in present-day New Mexico, in the United States, roughly 21 to 23 thousand years ago. The data come from a dry lakebed (or playa) at White Sands National Park that held water 19 to 36 thousand years ago. The lake bottom consists of gypsum and today plaster casts of human footprints are visible. According to the researchers, footprints were made when people walked through the shallow lake. They maintain that seeds imbedded in the footprints provide material for radiocarbon dating. The results can, arguably, pinpoint when the impressions were made. The reader can decider for him or herself if the researchers are making a persuasive argument. One huge problem with their conclusions, however, it that the Laurentide Ice Sheet prevented human migration into the interior of North America until about 14 thousand years ago.

At the end of the day, "pre-Clovis" represents a controversial archeological debate that fails to advance an understanding of Native American languages. In other words, linguists do not have a dog in this fight. We should view "pre-Clovis" as a moot point. Pre-Clovis arguments correlate the appearance of Clovis points with the first Native Americans. Very recent archeological and genetic evidence changes this long-standing assumption. Clovis simply represents the terminal end of a long mammoth hunter cultural tradition that ended about 11 thousand years ago in northern Eurasia and in the Americas.

9.5. The Polynesians.

Campbell (1997: 261-262) provides numerous "far-fetched" macro-family proposals that have surfaced in Native American historical linguistics: Amerindian and Basque; Na Dené and Mongolian; Mayan and Turkic; Quechua and Tungusic. His list of "far-fetched" proposals also includes Native American languages and Austronesian.

Unfortunately, despite the absence of genetic, linguistic, and archaeological evidence, the rumors of a prehistoric Polynesian migration into the Americas continue to circulate. The Polynesian rumor was once driven by a branch of archaeology called cephalometry, the study and measurement of the head. For example, a study from 1996 (Neves et al.) proposed that Polynesians were among the founding populations of the Americas based on craniometric measurement of 53 skulls. Another more famous example is "Kennewick man." He died about 8,000 years ago near Kennewick, Washington in the United States. For several years his remains were associated with Polynesian or Ainu ancestry based on skull measurements (e.g., Taylor, Smith and Southon 2001). However, as reported by Rasmussen et al. in 2015, Kennewick man belongs to Q1b-M3, a Native American lineage.

Genomic studies (Skoglund et al. 2015; Moreno-Mayar et al. 2018) have re-ignited the Native American-Polynesian rumor by reporting an Australasian component among Native American. The statistical modeling developed by Moreno-Mayar (2018) utilized an Andaman Islander with the P-P295 mutation as their Polynesian "proxy." It should be noted that P1-M45 mutations are found in South Asia (Gazi et al. 2013) and Island Southeast Asia (Karafet et al. 2015). However, the same mutation was part of the genetic inventory of Northern Eurasia during Marine Isotope Stage 3. The P1-M45 was recently found in Paleolithic remains from Siberia, the Yana-1 children who died 32 thousand years ago (Sikora et al. 2018). Rather than a Polynesian signal, Skoglund et al. (2015) and Moreno-Mayar et al. (2018) probably detected a prehistoric North Eurasian genetic signal among Native Americans, which is expected.

9.6. Solutreans.

Among the most dubious models surrounding the origins of Native Americas (Campbell (1997: 90-93) is that they came from the lost island of Atlantis. Close behind in the parade of dubious models is the Solutrean hypothesis, that Clovis is a continuation of the Solutrean cultural tradition that ended roughly 17 thousand years ago in Europe. Needless to say, this hypothesis is not consistent with mainstream archaeological opinion (see Straus, Meltzer, and Goebel 2005). Part of the problem is geographic distance. The lack of unequivocal Solutrean artifacts in the Americas is also problematic.

Despite the absence of archeological data, the Solutrean hypothesis has nevertheless resurfaced because of recent genetic data, the Q1b-L804 mutation. At this point the reader is directed to **Supplementary Figure 17.1**. As shown by the figure, Q1b-M3 and Q1b-L804 are sister clade mutations downstream from Q1b-M930. As previously discussed, Q1b-M3 is a Native American signature lineage. Q1b-L804, on the other hand, is found in northeastern Europe where it attains a very small frequency among the men of this region. Given the close phylogenetic relationship between Q1b-M3 and Q1b-L804, Wei et al. (2018) felt compelled to warn researchers that Q1b-L804 does not support the Solutrean hypothesis. Rather, the genetic data support a rapid diversification of the Q1b-L54 mutation around 16 thousand years ago in south central Siberia. Q1b-L804 and Q1b-L330 remained in the region. Q1b-M3 and Q1b-Z781 are the relics of Q1b-L54 diversification in the Americas. Taking this a step further, the sporadic appearance of haplogroups Q-M242 mutations in western Eurasia is explained by late Pleistocene expansions, founder effect, and genetic drift. This explanation is consistent with the data presented in Chapter 18 and the discussion of haplogroup R-M207.

Section 10. Conclusions for Haplogroup Q-M242.

The first humans that migrated into the Americas belonged to a cultural tradition that began 50 thousand years ago in Northern Eurasia, the mammoth hunters of the Upper Paleolithic. Eleven thousand years ago, with the onset of the Holocene, warmer weather caused a worldwide extinction of the mammoths. At this point, the Pleistocene mammoth-hunter tradition ended in Northern Eurasia and North America. Those that adapted to the Holocene and climate change in the New World became the Native American cultural tradition. This cultural tradition marks the beginning of linguistic diversity in the Americas.

The incredible diversity of indigenous languages in the Americas correlates well with the diversified subsistence strategy adopted by the Native Americans. They exploited regional resources. These resources brought opportunities and imposed constraints. In some areas, foraging and language evolved *in situ*. The Pacific Coast Indians are an example. However, some foraging cultures, such as the Athabaskans, migrated. In some areas, agriculture and language co-evolved. The co-evolution of agriculture and Quechuan in the Central Andes is an example. However, one finds evidence of language-farming expansions. Arawakan provides a solid example. Finally, some cultures, such the Numic peoples, abandoned agriculture and returned to foraging.

Amerind languages evolved directly from the Native American cultural tradition that formed at the beginning of the Holocene. Their evolution remained undisturbed from outside influence until 1492. Eskimo-Aleut, on the other hand, was shaped by contact with the indigenous people of northeastern Asia over the last five thousand years. Unfortunately, the paucity of genetic data prevents us from knowing more about the prehistory of the Eyak-Athabaskan language family. For example, it is difficult to utilize genetic data to defend or reject Greenberg's proposal that Eyak-Athabaskan stems from a second migration into the Americas. Additionally, geneflow between Alaskan Eskimos and Athabaskans remains an open question.

Chapter 18: Haplogroup R-M207.

Section 1. The Contemporary Distribution of Haplogroup R-M207.

Haplogroup R-M207 has two main divisions within its phylogeny, R1-M173 and R2-M479 (see, also, **Supplemental Figure 18.1**). The R2-M479 branch is found mostly among populations living in South Asia (see **Supplementary Table 18.1**). The R1a-M420 and R1b-343 mutations define the two main divisions within R1-M173. R1a-M420 mutations are found in Scandinavia, Eastern Europe, the Baltic Region, South Asia, Central Asia, and Northern Eurasia (**Supplementary Table 18.2**). R1b-M343 mutations are mostly found in Western Europe and the Sahel region of Africa (**Supplementary Table 18.3**).

Section 2. The Evolutionary History of Haplogroup R-M207.

Around 29 thousand years ago Marine Isotope Stage 3 ended. This marks the beginning of Marine Isotope Stage 2. Shortly thereafter, roughly 27 thousand years ago, the Ice Age glaciers reached their maximum southern extent across Eurasia. The literature refers to this event as the Last Glacial Maximum (see Clark et al. 2009 for a more detailed discussion). Within this climatological context, haplogroups R-M207 and Q-M242 evolved from the P1-M45 mutation roughly 33 thousand years ago (Poznik et al. 2016). See, also, **Supplementary Figure 1.1** from the first chapter.

Haplogroups R-M207 and Q-M242 represent the genetic signature of cold adapted populations who thrived and survived during the Last Glacial Maximum. Glaciation had driven these cultures southwards across Eurasia into what the literature describes as "refugia" (e.g., Gavashelishvili and Tarkhnishvili 2016). Ancient DNA data support for this position comes from the so-called "Mal'ta boy." He died about 24 thousand year ago near Lake Baikal in southern Siberia. As previously noted in Chapter 17, researchers place the child's Y-chromosome somewhere near the root of the R-M207 haplogroup, where haplogroups R-M207 and Q-M242 diverge from P1-M45.

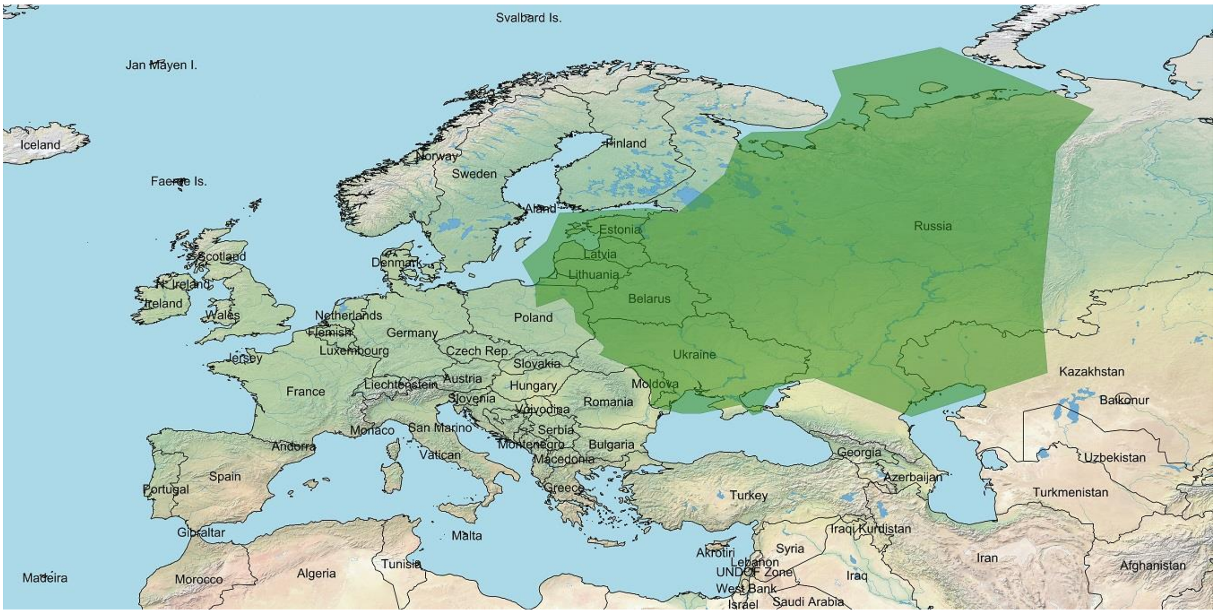
Turning now to the climatological record, at the time of the Last Glacial Maximum, Northern Eurasia and Eastern Europe were spared from the extreme glaciation that had occurred in Western Europe (e.g., Velichko et al. 2009; Binney et al. 2016). Across this vast expanse, cold and arid climatic conditions created a "hyperzonal open-type" biome that supported large mammal food resources needed by the cold adapted cultures. The position of this so-called "hyperzone" equates to around 50 degrees north for both regions (Velichko et al. 2009). According to Kuzmin (2008), because of the large mammal food resources that were supported by the hyperzone, human population density in this region remained relatively stable. Moreover, several factors contributed to successful cold weather adaptation for those living along the hyperzone: micro-blade tools, suitable dwellings and clothing, and the availability of bones for fuel.

Meanwhile the glacial ice sheet in Western Europe pushed much further south during the Last Glacial Maximum, almost to 40 degrees north, which equates roughly to the contemporary Franco-Spanish border. Kuzmin (2008) attributes greater glaciation in Western Europe to a shift in the Atlantic storm track that brought more moisture to the region. As a result, Scandinavia and much of Western Europe became depopulated and human populations retreated to the Iberian Peninsula. Interestingly, in contrast to refugia in South-Central Siberia and Eastern Europe, human population density within the Iberian refugium may well have suffered as the result of a decline in reindeer populations (Jochim et al. 1999; Morein 2008). Perhaps a population bottleneck explains the ultimate demise of C1a-V20 mutations that arrived in Europe during the Paleolithic. This, of course, leaves haplogroup I-M170 as the remaining founder lineage among contemporary Europeans. Taking this a step further, greater reproductive success during the Last Glacial Maximum stands as one of several factors that explain the ubiquitous presence of R-M173 mutations among the contemporary population of Eurasia.

Section 3. The Expansion and Diversification of R1-M173.

The R1-M173 mutation evolved about 30 thousand years (Poznik et al. 2016). The contemporary distribution of R2a-M124 mutations (see Section 10), along with the climatological and archeological evidence, as presented previously in Section 2, suggests that this occurred in southern Siberia. Then about 27 thousand years ago, R1a-M420 and R1b-M343 diverged from R1-M173 (Poznik et al. 2016). Underhill et al. (2015) suggest, based on analysis of contemporary Y-chromosome data, that the split occurred in the Middle East in the vicinity of Iran. However, archaeological and ancient DNA evidence places the diversification of R1-M173 on East European Plain in the vicinity of Kiev in the Ukraine. Archaeological support for this model stems from Abramova et al. and their 2001 report of Upper Paleolithic sites in the middle Dnieper River basin. These sites date from about 25 to 12 years thousand ago, and as such, they extend back in time to the Last Glacial Maximum. The ancient DNA support the distribution of R1a-M420 and R1b-M343 mutation in the Baltic region and Eastern Europe (see **Supplementary Tables 18.16** and **18.17**).

Figure 18.1. East European Plain.



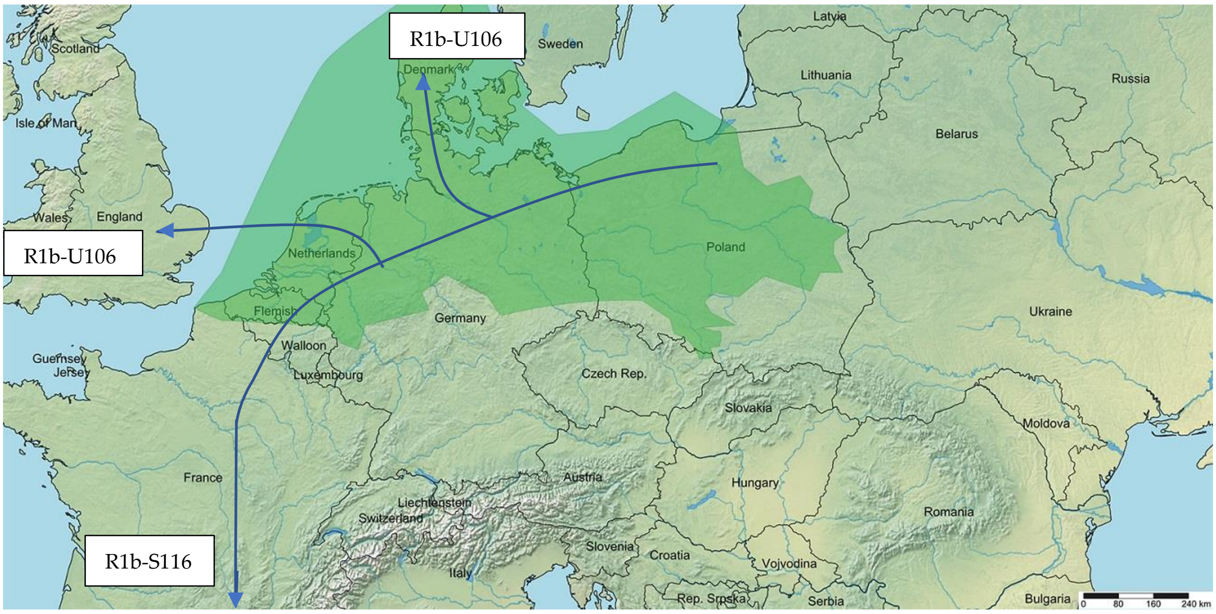
Section 4. Diversification of R1b-M343 on the East European Plain.

Dolukhanov (2009) provides a discussion of hunter-gatherer cultures at the time of the Paleolithic-Mesolithic transition of the Eastern European, about 12 thousand years ago. From a climatological perspective, this transition marks the beginning of the Holocene. Connected to this change in climate was the southward expansion of human populations on the East European Plain, from northeastern Europe to southeastern Europe. Southeastern European diversification of R1b-M343, on the other hand, is supported by the contemporary distribution of R1b-M73 (see **Supplementary Table 18.4**) as well as the evolutionary history of the R1b-V88 and R1b-CTS1078 mutations (see Sections 9 and 14 below).

Section 5. The Expansion of R1b-M269 into Western Europe.

R1b-M269 probably evolved in northeastern Europe. Around the onset of the Holocene, R1b-M269 expanded westward across the North European Plain into Western Europe (see Figure 18.2 below). The mutation was carried by reindeer hunters into Scandinavia, the British Isles, Iberia, and the Mediterranean. Archaeological support for this position partially stems from the discussion in Chapter 10. During the Last Glacial Maximum human populations in Western Europe had retreated to southern refugia including the Iberian Peninsula. Based on radio-carbon estimates, about 14 thousand years ago hunter-gatherers and reindeer in Western Europe began their expansion northwards (Housley et al. 1997). The glaciers slowly retreated. Tundra then appeared in previously glaciated area. The tundra eventually succumbed to the forests, and the tundra line gradually retreated to its present location above the Arctic Circle. The reindeer followed the retreating tundra line, and hunter-gatherers followed the reindeer (see Sommer et al. 2014 for additional details). By around 12 thousand years ago, the reindeer arrived in Denmark, and 9 thousand years ago, they disappeared into the arctic region of Scandinavia (Aaris-Sorensen et al. 2007).

Figure 18.2. North European Plain and the Diversification of R1b-L51 Mutations.



The archaeological evidence further suggests that around the beginning of the Holocene (12 thousand years ago) almost all of the mega-faunal food resources of northeastern Europe, such as mammoths, had disappeared (e.g., Puzachenko and Markova 2019). The last of the large herbivores was reindeer. Hunter-gatherers on the East European Plain, the Swiderian culture, were then drawn to Western Europe because of the abundance reindeer in this region and the demise of mega-faunal resources elsewhere in Eurasia (see Dolukhanov 2009).

Turning now to the genetic evidence, the data suggest that at the onset of the Holocene hunter-gatherers intercepted the northward migration of reindeer in Western Europe and Scandinavia along two different trajectories. As explained previously in Chapter 10, haplogroup I-M170 represents the genetic signature of the south to north trajectory. The R1b-L51 mutation, on the other hand, represents the genetic signature of the east to west trajectory. The R1b-L51 mutation is a downstream variant of the R1b-M269 mutation. According to Myres et al. (2011), this mutation evolved in Western Europe. The same study reports that during the early Mesolithic, R1b-U106 and R1b-S116 diverged from R1b-L51. R1b-U106 reflects genetic diversification of R1b-L51 on the North European Plain, the British Isles and Scandinavia. This position is supported by the contemporary distribution of R1b-U106 (see

Supplementary Table 18.5. The R1b-S116 mutation, on the other hand, reflects diversification of R1b-L51 on the Iberian Peninsula. Valverde et al. (2016) defend this position with the observation that the highest frequency of S116 mutations is found in Spain and Portugal.

Section 6. R1b-S116 and the Basque.

R1b-S116 has three informative downstream mutations: R1b-DF27, R1b-U152, and R1b-M529. Valverde et al. (2016) suggest based on their analysis of the genetic evidence that these three mutations diverged from R1b-S116 on the Iberian Peninsula around 12 thousand years ago, at the onset of the Holocene. They also take the position that during the Bronze Age, roughly 4,000 years ago, the R1b-DF27 mutation remained on the peninsula. R1b-U152, on the other hand, expanded eastwards onto the Italian Peninsula and then northwards through the Alps into Germany. Like R1b-U152, R1b-M529 also expanded out of the Iberian Peninsula. However, this mutation expanded along a different trajectory to the British Isles. This expansion model is supported by the contemporary distribution of all three mutations (see **Supplementary Tables 18.6**, **18.7**, and **18.8**).

One interesting observation from two genetic studies of Iberian populations, Valverde et al. (2016) and Solé-Morata et al. (2017), is that R1b-DF27 mutations attains a high frequency found among the Basque people. This a salient point for linguists because the Basque language is a potential relic of European linguistic diversity that predates the arrival of Indo-European languages (see discussion in Chapter 10). The R1b-S116 mutation, which arrived in the Basque region around the beginning of the Holocene, and the divergence of R1b-DF27 from R1b-S116 in Iberia, provide additional support for this position.

Section 7. The R1b-S116 Mutation and Celtic Languages.

Based on contemporary and ancient data, G-M201 mutations in Iberia stand as a genetic relic of the Neolithic transition in this region (see Chapter 8). The archeological record (see Martins et al. 2015) dates the arrival of agriculture and the Cardial Ware tradition on the Iberian Peninsula to around 7,500 years ago. This is based mostly on radio-carbon results taken from the remains of domesticated sheep and goats, an exceptionally reliable data source for delineating the Neolithic/Mesolithic transition. Their data further suggest that the Neolithic transition drove rapid population growth within the region. Taking this a step further, these data support the idea that population pressure drove an expansion of R1b-U152 and R1b-M529 mutations from the peninsula.

In contemporary Europe, Celtic languages are spoken in Ireland, the United Kingdom, and the Brittany region of France. *Ethnologue* (2018) places these languages into a single "Insular" branch. The Insular branch is further subdivided into Brythonic and Goidelic. Brythonic consists of Breton and Welsh. Irish and Scottish Gaelic form the Goidelic languages. In prehistoric Europe, Celtic languages had a much broader distribution extending from the Atlantic Ocean to Asia Minor. Historical Celtic languages not only included the Insular branch but also the extinct Continental Celtic branch of languages. Examples from the Continental branch include Celtiberian on the Iberian Peninsula, Gaulish in France, and Leponic in Northern Italy and Switzerland.

As noted previously in Section 6 (above), the R1b-DF27, R1b-M529, and R1b-U152 are downstream from the R1b-S116 mutation (see, also, **Supplementary Figure 18.1**). As shown by **Supplementary Tables 18.6**, **e 18.7**, and **18.8**, the available data for downstream variants of R1b-S116 are rather limited. Nevertheless, these data strongly link R1b-U152 with the historical distribution of Continental Celtic, and R1b-M529 with Insular Celtic.

A synthesis of several different data sources presented previously in this present chapter, and in Chapter 8, suggest farmers from southwestern Asia and proto-Indo-European languages co-expanded to the Iberian Peninsula during the Neolithic. In this region, proto-Celtic evolved from the language of the Neolithic farmers. Celtiberian represents later diversification of proto-Celtic within the same region. The other languages from the Continental branch, such as Gaulish and Lepontic, as well languages from the Insular branch, represent a co-expansion of Celtic-speaking farmers from Iberia and subsequent diversification of their language in other regions of Europe. This model of Celtic origins is strikingly similar to one that utilizes mitochondrial DNA (see McEvoy et al. 2004). Moreover, this model suggest that Celtic might have been shaped by Mesolithic non-Indo-European languages. A similar argument was made for Germanic in Chapter 10.

Section 8. Expansion of R1b-V88 into Mediterranean Europe and North Africa.

Interestingly, the oldest R1b-M343 sample comes from the Villabruna remains found in the Dolomite Mountains region of Northern Italy (see **Supplementary Table 18.15**). The remains are from a man who died around 14 thousand years ago. Researchers determined that he belongs to R1b1-L278, the ancestral mutation of R1b-M269 and R1b-V88. These remains provide strong support for a model of R1b-M343 variation in Europe that links the R1b-M269 and R1b-V88 mutations with the collapse of the Upper Paleolithic mammoth hunter tradition in Eurasia.

The highest frequencies of the R1b-V88 mutation among contemporary populations are found in the Sahel region of Africa, a zone that divides North Africa and Sub-Saharan Africa (see **Supplementary Table 18.9**). A synthesis of genetic, archaeological, and climatological perspectives suggest that the R1b-V88 mutation evolved in Eastern Europe during the Paleolithic. During the Mesolithic, the mutation expanded into Italy. A second Mesolithic expansion then carried the mutation from Sardinia into North Africa. Support for this position comes from a 2018 study (D’Atanasio et al.) that dates the evolution of R1b-V88 about to around 12 thousand years ago in Europe. Moreover, ancient DNA data pinpoints the evolution of R1b-V88 on the East European Plain. As shown by **Supplementary Table 18.17**, R1b-V88 remains from Eastern Europe are dated to around 11 thousand years ago.

D’Atanasio et al. (2018) take the position that R1b-V88 mutations in Africa stand as genetic relic of the North African Mesolithic and the so-called “humid phase.” As the reader may recall from the discussion in Chapter 2, the climate of northern Africa underwent a dramatic transformation about 10 thousand years ago. Holocene climate change brought rain to the region. The Sahara Desert became a savannah with numerous lakes and rivers. Within the complex system of rivers and lakes, hippos, crocodiles, and fish proliferated. The humans that hunted these animals thrived and survived.

The co-expansion of foragers and R1b-V88 could have potentially entered northern Africa via southwestern Asia. However, Kuper, and Kröpelin (2006) suggest that during the humid phase the Nile River was a marshland. As such, a Mesolithic expansion of R1b-V88 via the Middle East and northeastern Africa would have been difficult. On the other hand, palaeohydrological data gathered from satellite imagery (Drake et al. 2011) demonstrate the presence of river systems in North Africa during the humid phase. Taking this a step further, hunter-gatherers potentially made a water crossing from Sardinia to the North Africa. From Mediterranean coastline of North Africa, these rivers could have facilitated a southward migration to food resources in the vicinity of Lake Chad. Support for this model stems from D’Atanasio et al. (2018). According to the study, R1b-V88 mutations among contemporary Sardinians are older than African Rb-V88 mutations. Additionally, R1b-V88 was detected in a Sardinian sample from the Neolithic (see **Supplemental Table 18.17**), which supports the potential presence of the mutation among Mesolithic Sardinians.

Section 9. R1b-V88 and the Prehistory of African Languages.

For linguists, R1b-V88 mutations in North Africa help to decipher the prehistory of Chadic, a branch of the Afro-Asiatic language family. In a 2010 study, Cruciani et al. suggested that R1b-V88 is a genetic signature for Chadic languages based on the high frequency of the mutations among the Chadic-speaking populations of Africa. Such a position seems problematic for several reasons. Haber et al. (2016) disputed the conclusion from Cruciani et al. (2010) asserting that the oldest R1b-V88 mutations are found among the Laal-speakers, a language isolate of northern Africa. Data for contemporary African populations (see **Supplemental Table 18.9**) also show that R1b-V88 is found among speakers of Semitic and Berber languages, which are also branches of the Afro-Asiatic family. Additionally, R1b-V88 is well represented among several populations that speak either a non-Bantoid Niger-Congo languages or Nilo-Saharan languages. Finally, the E1b-M34, E1b-M81 and J1-M267 mutations, along with the archaeological data, suggest that Afro-Asiatic expanded into Africa from the Middle East during the Neolithic (see Chapters 5 and 11). R1b-V88, on the other hand, is a clearly a Mesolithic relic among African populations.

Section 10. Diversification of R1a-M420 Mutations.

As previously explained in Section 4, R1a-M420 and R1b-M343 diverged from R1-M173 about 25 thousand years ago on the East European Plain. Like R1b-M343 (see Section 3), the initial diversification of R1a-M420 variation occurred in northeastern Europe and southeastern Europe. R1a-M420 has two main downstream variants, R1a-Z282 and R1a-Z93 (see **Supplementary Figure 18.1**). Based on contemporary population data (see **Supplementary Table 18.11**) it appears as though the R1a-Z282 mutation represents diversification of R1a-M420 in northeastern Europe. Limited support for this position comes from ancient DNA data (see **Supplementary Table 18.16**). Contemporary data, on the other hand, suggest that R1a-Z93 evolved in Central Asia (see **Supplementary Table 18.14**).

Section 11. The R1a-Z282 Mutation and Slavic Languages.

R1a-Z282 has three informative downstream markers: R1a-Z284, R1a-M458, and R1a-M558. R1a-Z284 is confined almost exclusively to Scandinavia where it attains a frequency of around twenty percent among Norwegians (Underhill et al. 2015). Based on coalescent time estimates (Underhill et al. 2015) it appears as though the evolutionary history of R1a-Z284 appear to be similar to that of R1b-U106 (see Section 5). Specifically, both are genetic relics of Mesolithic reindeer hunters.

The R1a-M448 and R1a-M558 mutations are found in Eastern Europe. Based on contemporary data, R1a-M448 appears to have a higher frequency among West Slavic speakers (see **Supplementary Table 18.12**) and R1a-M558 seems to peak among East Slavic populations (see **Supplementary Table 18.13**). It should be emphasized that the contemporary distribution of R1a-M448 and R1a-M558 mutations are not linked with the historical Slavic expansion. Rather, language contact during historical times explains their contemporary distribution (see Chapter 10 for more details).

Section 12. R1a-Z93 and South Asian Languages.

As previously noted in Chapter 9 and the discussion of haplogroup H-M2713, the Central Asian *steppe nomad hypothesis* has surfaced in several studies that report Y-chromosome variation in the South Asian gene pool. Their analysis considers whether a Bronze Age invasion from Central Asia brought Indo-Aryan languages to South Asia. Consequently, R1a-Z93 becomes an especially significant marker for deciphering the linguistic prehistory of South Asia.

The geneticists have, for the most part, rejected any correlation between Y-chromosome variation in South Asia and a massive Bronze Age invasion of steppe nomads from Central Asia (e.g., Sahoo et al. 2006; Sengupta et al. 2006). Rather, the population history of the region has Paleolithic, Mesolithic, and Neolithic components (see Chapters 6, 8, 9, 11, 12, and 16). The R1a-Z93 mutation represents a Mesolithic component. Support for this conclusion is provided by Pamjav et al. (2012). According to the study, R1a-Z93 diverged from R1a-M420 about 10 thousand years ago, during the

Mesolithic.

Data previously presented above in this chapter links the evolutionary history of haplogroup R-M207 mutations with Holocene climate change. The mammoth hunting tradition of the Eurasian steppes collapsed because of warmer weather. The R1a-Z93 mutation stands as a Mesolithic genetic relic of hunter-gatherers who were driven into South Asia because of climate change in Central Asia or Eastern Europe. Archaeological support for this position comes from Virendra Nath Misra (2001), an expert in this field for South Asia. He reports that population density in India was low during the Upper Paleolithic. Arid and cold weather had limited the availability of food resources. The Mesolithic, however, brought monsoon rains. Increased moisture produced more food resources that ultimately drove higher population density.

Section 13. The R2-M479 Mutation and South Asian Languages.

As noted previously in Section 1, R1-M173 and R2-M479 form the two main downstream divisions of the R-M207 haplogroup. Data from Poznik et al. (2016) suggest that the R2-M479 mutation evolved roughly 30 thousand years ago, around the time of the Last Glacial Maximum. As suggested in Section 2, this appears to have occurred in southern Siberia.

Contemporary data for the R2-M479 mutation consists almost entirely of frequency results for the R2a-M124 downstream mutation. As shown by **Supplementary Table 18.1**, almost all the reported data for R2a-M124 comes from South Asia. Sengupta et al (2006) suggest that about 9 percent of Indian and 7 percent of Pakistani males have the R2a-M124 mutation. Additionally, the mutation attains a moderate frequency among Indo-Aryan and Dravidian-speaking populations (see **Supplementary Table 18.1**).

Like the R1a-Z93 mutation, R2-M479 variation in South Asia also stands as a Mesolithic component among the populations of this region. Climate change also drove these mutations and hunter-gatherers into the region. Support for this position is provided by ancient DNA. As shown by **Supplementary Table 18.18**, R2a-M124 was extracted from a 10-thousand-year-old sample found at Ganj Dareh in northwestern Iran.

Section 14. Problematic Palaeogenomic Modeling of Indo-European.

"Palaeogenomic modeling" attempts to model human population history by employing statistical analysis of ancient DNA, and more specifically, autosomal DNA markers that are inherited from both parents. Statistical methods are utilized in order to overcome the problem of recombination. This "reshuffling" of genetic traits can distort and erase evolutionary relations that are needed to decipher human population history. Of course, Y-chromosome data avoid this problem as they are gathered from a non-recombining region of the human genome (see Chapter 1).

An especially problematic palaeogenomic model of Indo-European languages in Europe was published by Haak et al. in 2015. Published in the journal *Nature*, the study endorses the *steppe nomad hypothesis* of Indo-European language origins based on statistical analysis of ancient autosomal markers. Moreover, the report asserts that the steppe nomads replaced 75 percent of the pre-existing farmer genes in Central Europe. Such a conclusion is flagrantly inconsistent with the archaeological record. Such a conclusion is also inconsistent with the Y-chromosome data. According to the ancient DNA data, the genetic signature of these nomads is the R1b-CTS1078 mutation (see **Supplemental Table 18.19**). The contemporary data report a virtual absence of the R1b-CTS1078 mutation among contemporary European populations (see **Supplementary Table 18.10**). If the genome of contemporary Europeans contains a large steppe nomad component from the Bronze Age, as reported by Haak et al. (2015), the R1b-CTS1078 mutation should exhibit a much higher frequency among the population of the European continent, which is clearly not the case.

Despite inconsistencies with the archaeological and genetic data, the position taken by Haak et al. (2015) continues to circulate and has been endorsed by several subsequent palaeogenomic studies published in peer-reviewed science journals (e.g., Allentoft et al. 2015; Jones et al. 2015; Cassidy et al. 2016; Jones et al. 2017; Olalde et al. 2018; Narasimhan et al. 2019; Egfjord et al. 2021). For linguists, the choice is clear. We can endorse a model of Indo-European origins based on statistical analysis of a small dataset. Alternatively, we can gravitate towards a triangulated Y-chromosome-based perspective that utilizes a synthesis of genetic, linguistic, archeological, and climate data that are drawn from a much larger data set.

Section 15. Conclusions for Haplogroup R-M207.

According to the triangulated Y-chromosome-based perspective, the contemporary distribution of R-M207 stands as a genetic relic of demographic processes that began at the end of the Last Glacial Maximum: expansions from Ice Age refugia; the demise of mega-fauna food resources on the Eurasian steppes; and rapid population growth associated with the Neolithic transition. For linguists, haplogroup R-M207 and its downstream variants provide especially useful makers for deciphering the prehistory of the Indo-European, Dravidian, Afro-Asiatic, and Nilo-Saharan language families as well as the Basque and Laal language isolates. One interesting observation from the genetic and archeological data is that language contact appears to partially explain the prehistory of Indo-European and Afro-Asiatic language families. R-M207 was already in South Asia and Europe when Indo-European-speaking farmers arrived in these regions. Similarly, R-M207 was in northern Africa when Afro-Asiatic-speaking farmers arrived.

The data from this chapter and that from the previous (Chapter 17) present another interesting observation for the anthologists. The mammoth hunter tradition of the Upper Paleolithic appears not to have exited the stage quietly. Rather, hunter-gatherers in Eurasia and the Americas fought Holocene climate change to the bitter end and then adapted.

Chapter 19: Recommendations, Observations, and Future Research.

Section 1. Recommendations.

Since Y-chromosome data are useful for modeling the prehistory of language, it would be in the best interest of linguists to encourage efforts that gather data from populations that represent the full spectrum of linguistic diversity. Currently the amount of contemporary Y-chromosome data varies greatly from one region to the next. For example, European populations have been studied extensively. On the other hand, comparatively little data exists for Sub-Saharan Africa, Southeast Asia (especially Myanmar, Vietnam, and Malaysia), the highlands region of New Guinea, and the indigenous peoples of North America (especially Alaska). Additionally, most of the data for aboriginal Australians has been taken from government databases that do not track group affiliation.

In some cases, the paucity of data for a region may reflect the availability of funding that is available for population studies. European countries, for example, have the financial resources for these studies, whereas a country like Papua New Guinea may lack the resources. Moreover, it should also be emphasized that Native North American and aboriginal Australian populations have generally refused to participate in genetic studies because of an historical distrust of Europeans. These groups certainly represent a key component in understanding the evolution of language. Hopefully we can build alliances with them in the future.

Section 2. Miscellaneous Observations.

2.1. Overview.

The non-recombining region of the human acts as a “trap” which has successfully captured important demographic milestones that mark the evolutionary history of *Homo sapiens*. Why this occurs is not completely understood as Y-chromosome mutations are selectively neutral, meaning they confer they do not affect reproductive success. Nevertheless, Y-chromosome mutations consistently record the reproductive success of our species, which, in turn, greatly enhance our understanding of the prehistory of language.

2.2. Language Shift and Y-Chromosome Diversity.

We inherit the mother tongue and genes from our parents. Language and genetics maintain a robust correlation despite the observed phenomenon of language shift. Y-chromosome evidence of language shift comes from atypical Y-chromosome mutations found in a population. For example, the N1a-M46 mutation consistently appears in Uralic-speaking populations. The unexpected appearance of this mutation among Lithuanians and Latvians is explained by language shift from Uralic to Baltic. The absence of N1a-M46 among Hungarians is explained by language shift from an unknown language to Uralic.

2.3. Linguistically Informative Mutations.

Y-chromosome data explain, at least partially, the contemporary pattern of global linguistic diversity. Chapters 2 to 18 explore the prehistory of language from a Y-chromosome perspective. The data suggest that triangulated Y-chromosome-based models of language prehistory are highly reliable. The initial step in building such models is to identify informative Y-Chromosome mutations among contemporary populations for which language has a strong ethnic component. "Informative" generally means that a mutation has a moderate to high frequency (> 10%) among speakers of a specific language family. This monograph has identified 110 linguistically informative Y-chromosome mutations: A1b-V50, A1b-M51, A1b-M13, B2a-M150, B2b-M112, D1a-F6251, D1b-M55, D1c-Y34638, E1a-M33, E1b-M81, E1b-M293, E1b-V13, E1b-V22, E1b-V32, E1b-M34, E1b-U174, E1b-U175, E2a-M41, C1a-M356, C1a-M8, C1b-M38 (x M208), C1b-M208, C2b-M48, C2b-F1918, C2b-F7171, C2b-F5484, C2c-CTS2657, G1-M285, G2a-L91, G2a-L497, G2a-L406, G2a-M527, G2b-M377, G2a-P16, G2a-U1, H1a-M69, I1-M253, I2a-M423, I2a-M26, I2a-M223, J1-M267, J1a-P58, J2a-M410, J2a-M67, J2a-M47, J2a-M68, J2a-M319, J2a-L24, J2b-M12, J2b-M24, L-M20, L1a-M27, L1a-M357, L1b-M317, T1a-M70, M-P256, S-B254, M1a-P34, S1a-M254, S1a-P308, M1a-P87, M2-M353, M3-P117, S1a-P79, N1b-F2930, N1a-P43, N1a-M46, N1a-B211, N1a-Z1936 N1a-M2019, N1a-VL29, N1a-F4205, N1a-B202, N1a-M128, N1a-B187, N1a-B479, O2a-002611, O2a-M117, O2a-F114, O1a-M307, O1a-M110, O2a-P164, O1b-M95, O2a-M7, O2a-M133, O1b-SRY465, O1b-47z, O1b-F1252, O1b-M111, O1b-M1283, O1a-M119, Q1b-M3, Q1b-Z780, C2b-P39, Q1a-M25, Q1a-F746, R2a-M124, R1a-M420, R1a-Z282, R1a-M458, R1a-M558, R1a-Z93, R1b-M343, R1b-M73, R1b-U106, R1b-DF27, R1b-U152, R1b-M529, R1b-V88, and R1b-CTS1078

The next step in the model building process is to explain why a mutation attains a significant frequency. To resolve this question, data is extrapolated from phylogenetic relationships, ancient DNA, language relationships, the archaeological record, the paleo-climatological record, and other relevant sources such as other genetic marker perspectives. The N1a-M46 mutation, for example, attains a significant frequency among Uralic-speaking populations because the domestication of reindeer improved reproductive success among a population living on the Taymyr Peninsula about 8,000 years ago.

2.4. Prehistoric Population Expansions and Climate Change.

Prehistoric population expansions help to explain the contemporary pattern of language variation. Major expansions include the out-of-Africa exodus during Marine Isotope Stage 5, the colonization of East Asia and Europe during Marine Isotope Stage 3, expansions from Ice Age refugia into the Americas during the late Pleistocene, and Neolithic agricultural expansions that occurred independently in several regions of the world. Interestingly, prehistoric human expansions were motivated, in part, by climate change. For example, less precipitation drove the out-of-Africa exodus. Warmer climatic conditions drove the human colonization of Europe, East Asia, and Australia during Marine Isotope Stage 3. Late Pleistocene deglaciation drove the human settlement of the Americas. The return of monsoon rain during the Holocene drove the East Asian Neolithic.

2.5. Reproductive Success and Language.

Agriculture has drastically improved the reproductive success of *Homo sapiens*. From an evolutionary perspective, this survival strategy supports far more people per square kilometer than foraging. Rice, for example, supports over a billion people in East Asia. For linguists, agriculture is a salient point because the Neolithic revolution drove rapid population growth. As a result, language and farmers co-expanded in several regions of the world. These language-farmer expansions, in turn, partially explain the evolutionary history of the following language families: Indo-European, Niger-Congo, Afro-Asiatic, Uralic, Sino-Tibetan, Austro-Asiatic, Dravidian, Austronesian, Trans-New Guinea, and Arawak (Maipurean).

2.6. Human Evolutionary Adaptations and Language.

Human evolutionary adaptations help to explain linguistic diversity. Admixture between Neanderthals and *Homo sapiens* may have strengthened the human immune system. The success of Tibeto-Burman languages stems from an evolutionary adaptation that enables Tibetans to utilize the depleted oxygen level found on the Tibetan Plateau. They can overcome hypoxia and altitude sickness, significant health risks among those that inhabit this region. Similarly, Austronesians were able to farm the coastal areas of New Guinea because of an evolutionary adaptation that made them resistant to tropical splenomegaly syndrome, a massive and fatal enlargement of the spleen that occurs as the result of chronic exposure to malaria.

2.7. Language Contact Theory.

From a triangulated Y-chromosome perspective, language contact theory provides an especially robust explanation for the global pattern of contemporary language variation. For example, the evolutionary history of Germanic languages involves language contact between the Mesolithic populations of Scandinavia and the Neolithic populations of Central Europe. The evolution of Indo-Aryan also involves language contact between Neolithic farmers and Mesolithic hunter-gatherers. The Pygmies of the central African rainforest adopted the Niger-Congo languages of Bantu farmers. Papuans adopted the languages of Austronesian farmers and then expanded eastwards across the Pacific. In North and East Africa, during prehistoric times, Nilo-Saharan speaking populations shifted to Afro-Asiatic, and Afro-Asiatic populations shifted to Nilo-Saharan. The story of Austronesian languages in western Indonesia and Malaysia entails language shift from Austro-Asiatic to Austronesian. Finally, language contact best explains similarities found among the so-called Altaic languages.

2.8. Trajectories of Prehistoric Language Evolution.

Language entered the historical record along five different trajectories. The first trajectory involves a co-expansion of early farming and language, similar to Bellwood’s *early farming dispersal hypothesis*. Indo-European, Niger-Congo, Uralic, Sino-Tibetan, Austronesian, Austro-Asiatic, and Maipurean are examples. The second trajectory involves the *in-situ* co-evolution of agriculture and language. Korean, Japanese, Nahuatl, Mayan, and Quechuan are examples. The third trajectory involves a co-expansion of hunter-gatherers and language. Eyak-Athabaskan is a good example. The fourth trajectory involves *in-situ* co-evolution of hunter-gatherers and language. Good examples come from North America and include Eskimo-Aleut, Tsimshian, Wakashan, and Salish. The fifth and final trajectory involves reversion from agriculture to foraging, a rarely observed phenomenon. Examples include Finnic languages and Numic.

Section 3. Unresolved Research Questions for the Future.

3.1. Introduction.

It was never my intent to create a new specialization in the field of linguistics. I recommend strongly that early career linguists specialize in one of the traditional sub-disciplines of linguistics, such as historical linguistics, or semantics, or morphology. We are linguists and not geneticists or archeologists or climatologists. Nevertheless, language is the repository of culture. Those who gravitate towards linguists recognize that our discipline can be a forum for exploring the vast spectrum of cultural diversity, both past and present. Triangulated Y-chromosome-based modeling represents a vehicle for delivering this package. Moreover, archeological, historical, anthropological, genetic, and climate data can offer useful perspectives for interpreting the linguistic data. For those willing to take a risk, triangulated Y-chromosome-based modeling of language prehistory offers an exciting opportunity to set sail into an unexplored linguistic frontier. So if you want to change the world, be a linguist who dares to be different.

3.2. The Comparative Method and Language Prehistory.

One question involves the comparative method and the potential contribution provided by this linguistic tool for elucidating the prehistory of language. The linguist Lyle Campbell, for example, tends to employ a conservative application of the comparative method in an effort to classify language diversity. The linguist Robert Beekes, on the other hand, employs the comparative method to reconstruct the culture of "Indo-Europeans," a people who were (apparently) especially good poets although this has never been documented either by the archaeological or historical record. Perhaps the comparative method is robust tool for language classification and fails miserably as a tool for cultural reconstruction.

3.3. Non-Linguistic Data and Language Classification.

Should we utilize non-linguistic data to classify languages? Controversial classifications such as Nilo-Saharan and Niger-Congo seem more robust when considers the genetic and anthropological data. Nilo-Saharan correlates well with the desertification of the Sahara and the evolution of cattle herding as a subsistence strategy in East Africa. Niger-Congo correlates well with the expansion of land agriculture from West-Central Africa. On the other hand, classifications such as Sino-Tibetan seem less plausible when considers non-linguistic evidence. From an anthropological perspective, for example, Chinese languages are linked to the success of rice agriculture in East Asia. Tibeto-Burman, on the other hand, evolved from the success of barley agriculture on the Tibetan and an evolutionary adaption among the Tibetan farmers that enables them to survive hypoxia. Additionally, non-linguistic perspectives favor the traditional "binary" model of Uralic.

3.4. Neolithic Revolution and the Leveling of Linguistic Diversity.

The special relationship between language variation and agriculture raises an interesting research question. Is there an inverse relationship between agriculture and language? In other words, did the Neolithic revolution level linguistic diversity? This question arises from the observation that linguistic diversity in South America has been difficult to classify. Compared to Eurasia and Africa, many of the South American languages are listed by *Ethnologue* as isolates or as unclassified. Additionally, linguistic diversity in South America consists of numerous small language families, whereas linguistic diversity in the Old World consists of comparatively fewer language families which in many cases, consist of hundreds of languages. As such, one could argue that linguistic distance is greater for New World languages than for Old World languages. Extending this argument further, this dichotomy may reflect that agriculture was practiced less intensively in prehistoric South America. On the other hand, the classification of indigenous language diversity in South America may reflect the availability of resources for historical linguistics. Alternatively, European colonization may have erased large sections of the linguistic map and as such, this has obscured linguistic relationships that facilitate language classification.

Section 4. Final Thoughts.

Interest in the prehistory of language has circulated within linguistic debate since the founding of our discipline over two hundred years ago. Nevertheless, among contemporary linguists some believe that the question of language prehistory is far too speculative, that it defies empirical analysis. This approach to linguistic research is problematic because the past explains the present. The more we know about the prehistory of language, the more we know about contemporary languages. Towards this goal some linguists have attempted to model the prehistory of language by using linguistic tools. These attempts, however, have rendered models of language that are sometimes clearly implausible, such as the correlation between the Basque isolate and the languages of the Caucasus region. Linguistic tools are also limited in achieving time depth. Triangulated Y-chromosome-based modeling of language prehistory yields desperately needed empirical models of language prehistory that are highly reliable. Moreover, we can drill much deeper into the prehistory of language.

Index of Languages and Cultures

Licensing Information for Figures

Figure 1.1. The Structure of DNA.

Source: Wikipedia and Madprime.
Accessed: August 3, 2021.
Page URL: https://commons.wikimedia.org/wiki/File:DNA_chemical_structure.svg
Licensing Information: CC BY-SA 3.0.

Figure 2.1. Linguistic Map of Africa.

Source: Wikipedia and Mark Dingemanse.
Accessed: August 16, 2021.
Page URL: https://commons.wikimedia.org/wiki/File:Khoi-San_NL.png
Notes: Original file was modified using different colors for language areas and translating Dutch into English.
Licensing Information: Creative Commons Attribution 2.0 Generic License.

Figure 4.3. Two Ainu Men, Tokoro Village, Hokkaido Island, Japan, between 1885 and 1895.
Source: Smithsonian Institution
Accessed: October 28, 2021.
URL: https://edan.si.edu/slideshow/viewer/?eadrefid=NAA.PhotoLot.77-38
Licensing Information: In the Public Domain. Romyn Hitchcock photographs of Ainu people, Japan, and Korea (NMNH-04753000b). National Anthropological Archives, Smithsonian Institution.

Figure 4.4. Andaman Islanders about 1911.
Source: Smithsonian Institution.
Accessed: October 7, 2021.
URL: https://sova.si.edu/details/NAA.PhotoLot.97#ref4676
Licensing Information: In the Public Domain. Lot 97 DOE Asia: Asiatic Islands: Andamans: People (NAA INV.04422400), National Anthropological Archives, Smithsonian Institution.

Figure 5.1. The Fertile Crescent 7500 BC.
Source: Wikipedia and Bjoertvedt.
Accessed: August 25, 2021
Page URL: https://commons.wikimedia.org/wiki/File:Fertile_crescent_Neolithic_B_circa_7500_BC.jpg
Notes: Red squares represent Neolithic sites, and pre-Neolithic sites are represented by black squares.
Licensing Information: Creative Commons Attribution-Share Alike 3.0 Unported License.

Figure 6.1. Sunda and Sahul.
Source: Wikipedia and Maximilian Dörrbecker (Chumwa).
Accessed: August 28, 2021
Page URL: https://commons.wikimedia.org/wiki/File:Map_of_Sunda_and_Sahul.png
Licensing Information: Creative Commons Attribution-Share Alike 3.0 Unported License.

Figure 6.2. Island Southeast Asia, Australia, and Oceania.
Source: United States Government and Central Intelligence Agency.
Accessed: August 28, 2021
Page URL: https://www.cia.gov/the-world-factbook/static/97fb508232ccc526320d30d41033ca09/oceania_phy.jpg
Licensing Information: In the Public Domain.

Figure 7.2. Ice Sheets, Tundra, and Steppe at the Last Glacial Maximum.

Source: Wikipedia and Fährtenleser
Accessed: August 31, 2021
Page URL:
https://de.wikipedia.org/wiki/Letzteiszeitliches_Maximum#/media/Datei:Vegetationszonen_18.000_BC.png
Notes: I added text to the map: "Asia," Europe," and "North America."
Licensing Information: Creative Commons Attribution-Share Alike 4.0 International license.

Figure 8.2. European Neolithic.
Source: Wikipedia
Accessed: September 1, 2021
Page URL: https://commons.wikimedia.org/wiki/File:Expansion_n%C3%A9olithique.png
Notes: I replaced French text with English text.
Licensing Information: In the Public Domain.

Figure 9.1. Language Families in South Asia.
Source: University of Buffalo, Department of Linguistics.
Accessed: October 1, 2021.
URL: https://www.acsu.buffalo.edu/~dryer/family_maps.htm
Licensing Information: Fair Use Exception of the US Copyright Act.

Figure 9.2. Hittite Cuneiform Tablet.
Source: Wikipedia and Mx. Granger
Accessed: September 13, 2021
Page URL: https://upload.wikimedia.org/wikipedia/commons/thumb/a/a7/Hittite_Cuneiform_Tablet-_Legal_Deposition%28%3F%29.jpg/1024px-Hittite_Cuneiform_Tablet-_Legal_Deposition%28%3F%29.jpg
Licensing Information: Creative Commons CC0 1.0 Universal Public Domain Dedication.

Figure 10.2. Photo of a Norwegian Reindeer.
Source: Wikipedia and Are G Nilsen.
Accessed: October 30, 2021
Page URL:
https://en.wikipedia.org/wiki/Reindeer#/media/File:Reinbukken_p%C3%A5_frisk_gr%C3%B8nt_beite._-_panoramio.jpg
Licensing Information: Creative Commons Attribution-Share Alike 3.0 Unported license.

Figure 10.3. Tundra.
Source: Wikipedia and Dr. Andreas Hugentobler
Accessed: September 15, 2021
Page URL: https://commons.wikimedia.org/wiki/File:Tundra_in_Siberia.jpg
Licensing Information: Creative Commons Attribution 2.0 Germany.

Figure 11.1. Manishtusu Obelisk and Akkadian Inscription.
Source: Wikipedia and Mbzt.
Accessed: September 16, 2021
Page URL:
https://commons.wikimedia.org/wiki/File:P1050578_Louvre_Ob%C3%A9lisque_de_Manishtusu_d%C3%A9tail_rwk.JPG
Licensing Information: GNU Free Documentation License, Version 1.2 and Creative Commons Attribution 3.0 Unported license.

Figure 11.2 Distribution of Afro-Asiatic and its Primary Branches 1000-2000 Years Ago.
Source: Wikipedia and Tussna.
Accessed: September 16, 2021
Page URL: https://commons.wikimedia.org/wiki/File:Afro-Asiatic_languages_-_Karte_der_Afroasiatischen_Sprachen.png#mw-jump-to-license
Notes: Image was adapted by removing German written test, by removing legend, and adding English-written labels for language branches.
Licensing Information: GNU Free Documentation License, Version 1.2 and Creative Commons Attribution-Share Alike 4.0 International license.

Figure 13.1. Uluru (Ayers Rock), Northern Territory, Australia.
Source: Wikipedia and Stuart Edwards
Accessed: September 22, 2021
Page URL: https://en.wikipedia.org/wiki/Uluru#/media/File:Uluru_Panorama.jpg
Licensing Information: Own work, copyleft: Multi-license with GFDL and Creative Commons CC-BY-SA-2.5 and older versions (2.0 and 1.0).

Figure 17.2. Bølling-Allerød Interstadial.
Source: Wikipedia; Platt, D. et al. 2017. "Mapping Post-Glacial expansions: the peopling of Southwest Asia. *Scientific Reports* 7**:** 40338. https://www.nature.com/articles/srep40338
Accessed: September 26, 2021
URLs: https://commons.wikimedia.org/wiki/File:Evolution_of_temperature_in_the_Post-Glacial_period_according_to_Greenland_ice_cores.jpg
https://www.nature.com/articles/srep40338
Licensing Information: Creative Commons 4.0 license.

Figure 17.4. Teosinte, Teosinte-Maize Hybrid, and Maize.
Source: Wikipedia and John Doebley
Accessed: September 27, 2021
URL: https://commons.wikimedia.org/wiki/File:Maize-teosinte.jpg
Licensing Information: Creative Commons Attribution 3.0 Unported license.

Figure 17.5. Lamas.
Source: Wikipedia and Kallerna
Accessed: September 27, 2021.
URL: https://en.wikipedia.org/wiki/File:Lama_glama_Laguna_Colorada_2.jpg
Licensing Information: Creative Commons Attribution-Share Alike 4.0 International license.

Figure 17.7. Amazon River Basin.
Source: Wikipedia and Kmusser.
Accessed: September 27, 2021.
URL: **https://commons.wikimedia.org/wiki/File:Amazonrivermap.png**
Licensing Information: Creative Commons Attribution-Share Alike 3.0 Unported license.

Figure 17.9. Distribution of Athabaskan Languages.
Source: National Park Service.
Accessed: September 28, 2021.
URL: https://www.nps.gov/chir/learn/historyculture/pre-apache-wars.htm
Licensing Information: In the public domain.

Figure 17.10. Japan, Russia, and Alaska.
Source: University of Texas and the Perry-Castañeda Library Map Collection.
Accessed: September 28, 2021
URL: https://maps.lib.utexas.edu/maps/asia.html
Licensing Information: In the public domain.
Notes: Created as smaller map, deleted original text, and added larger text.

The remaining figures appearing in the monograph belong to the author himself and retain the same license as the monograph itself. Maps were prepared using SimpleMappr, an open-source online tool. Shorthouse, David P. 2010. SimpleMappr, an online tool to produce publication-quality point maps. [Retrieved from https://www.simplemappr.net. Accessed November 19, 2021].

Bibliography.

Aaris-Sorensen, Kim et al. 2007. “The Scandinavian reindeer (*Rangifer tarandus L.*) after the last glacial maximum: time, seasonality and human exploitation.” *Journal of Archaeological Science* 34(6): 914-923.

Abi-Rached, Laurent et al. 2011. “The shaping of modern human immune systems by multiregional admixture with archaic humans.” *Science* 334 (6052): 89-94.

Abramova et al. 2001. “The age of Upper Paleolithic sites in the middle Dnieper River basin of Eastern Europe.” *Radiocarbon* 43(2b): 1077-1084.

Abu-Amero, Khaled K. et al. 2009. “Saudi Arabian Y-chromosome diversity and its relationship with nearby regions.” *BioMed Central Genetics* 10:59.

Adcock, Gregory J. et al. 2001. “Mitochondrial DNA sequences in ancient Australians: Implications for modern human origins.” *Proceedings of the National Academy of Sciences of the United States of America* 98(2): 537–542.

Ahn, Sung-Mo 2010. “The emergence of rice agriculture in Korea: archaeobotanical perspectives.” *Archaeological and Anthropological* Sciences 2: 89-98.

Aikio, Ante 2020. “Loanwords from unattested Nordic source forms in Saami”. *Finnisch-Ugrische Forschungen* 2020 (65), 5–24.

Aiko, Ante in Press. Proto-Uralic. — To appear in: The Oxford Guide to the Uralic Languages. Edited by Marianne Bakró-Nagy, Johanna Laakso, and Elena Skribnik. Oxford University Press. Retrieved from Academia on February 14, 2021. https://www.academia.edu/40193033/Proto_Uralic

Allen, Bryan 1992. “The geography of Papua New Guinea.” In: *Human Biology in Papua New Guinea: the Small Cosmos*. Edited by Robert D. Attenborough and Michael P. Alpers. Oxford; New York; Tokyo; Melbourne: Clarendon Press, pp. 36-66.

Allen, Jim and James F. O'Connell 2008. “Getting from Sunda to Sahul.” In*: Islands of Inquiry: Colonisation, Seafaring and the Archaeology of Maritime Landscapes*. Edited by Geoffrey Clark, Foss Leach, and Sue O’Connor. Canberra: ANU Press, pp. 31-46.

Allentoft, M. E. et al. 2015. “Population genomics of Bronze Age Eurasia.” *Nature* 522: 167-172.

Ammerman, A.J. and L.L Cavalli-Sforza 1984. *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton: Princeton University Press.

Amos, William 2020. “Signals interpreted as archaic introgression appear to be driven primarily by faster evolution in Africa.” *Royal Society Open Science* 7: 191900.

Amos, William 2021. “Correlated and geographically predictable Neanderthal and Denisovan legacies are difficult to reconcile with a simple model based on inter-breeding.” *Royal Society Open Science* 8: 201229.

Anikovich, M.V. et al. 2007. “Early Upper Paleolithic in Eastern Europe and implications for the dispersal of modern humans.” *Science* 315: 223-226.

Anthony, David W. 2007. *The Horse, the Wheel, and Language, How Bronze-Age Riders from the Eurasian Steppes Shaped the Modern World.* Princeton, N.J.; Woodstock: Princeton University Press.

Anthony, David W. and Don Ringe 2015. "The Indo-European homeland from linguistic and archaeological perspectives." *Annual Review of Linguistics* 1: 199-219.

Anthony, David W. 2017. "Archaeology and Language: Why archaeologists care about the Indo-European problem." In: *European Archaeology as Anthropology: Essays in Memory of Bernard Wailes*. Edited by Pam J. Crabtree and Peter Bogucki. University of Pennsylvania Press, pp. 39-70.

Arbuckle, Benjamin S. and Emily L. Hammer 2019. "The rise of pastoralism in the ancient Near East." *Journal of Archaeological Research* 27: 391-449.

Arredi, Barbara et al. 2004. "A predominately Neolithic origin for Y-chromosomal DNA variation in North Africa." *American Journal of Human Genetics* 75: 338-345.

Arrow, K.J., Panosian C., and Gelband H. (editors) 2004. *Saving Lives, Buying Time: Economics of Malaria Drugs in an Age of Resistance.* Washington (DC): Institute of Medicine (US) Committee on the Economics of Antimalarial Drugs. National Academies Press, pp. 125-135. https://www.ncbi.nlm.nih.gov/books/NBK215638/

Arunkumar, Ganesh Prasad et al. 2012. "Population differentiation of southern Indian male lineages correlates with agricultural expansions predating the caste system." *Public Library of Science One* 7(11): e50269.

Arunkumar, Ganesh Prasad et al. 2015. "A late Neolithic expansion of Y chromosomal haplogroup O2a1-M95 from east to west: late Neolithic expansion of O2a1-M95." *Journal of Systematics and Evolution* 53(6): 546-560.

Arvidsson, Stefan 2006. *Aryan Idols: Indo-European Mythology as Ideology and Science.* Chicago: University of Chicago Press.

Bahuchet, Serge 2012. "Changing language, remaining Pygmy." *Human Biology* 84 (1):11-43.

Bailey, Geoff et al 2007. "Coastal prehistory in the Southern Red Sea Basin, underwater archaeology and the Farasan Island." *Proceedings of the Seminar for Arabian Studies* 37: 1-16

Bailey, Robert C. et al. 1989. "Hunting and gathering in tropical rain forest: is it possible?" *American Anthropologist* 91: 59-82.

Bajić, Vladamir et al. 2018. "Genetic structure and sex-biased gene flow in the history of southern African populations." *American Journal of Physical Anthropology* 167(3): 656-671.

Balanovsky, O. et al. 2008. "Two sources of the Russian patrilineal heritage in their Eurasian context." *American Journal of Human Genetics* 82: 236-250.

Balanovsky et al. 2011. "Parallel evolution of genes and languages in the Caucasus region." *Molecular Biology and Evolution* 28(10): 2905-2920.

Balanovsky, Oleg et 2015. "Deep phylogenetic analysis of haplogroup G1 provides estimates of SNP and STR mutation rates on the human Y chromosome and reveals migrations of Iranic speakers." *Public Library of Science One* 10(4): e0122968.

Balanovsky, O. 2017a. "Toward a consensus on SNP and STR mutation rates on the human Y-chromosome." *Human Genetics* 136:575-590.

Balanovsky, O. et al. 2017b. "Genetic differentiation between upland and lowland populations shapes the Y chromosomal landscape of West Asia." *Human Genetics* 136: 437-450.

Balaresque, Patricia et al. 2010. "A predominantly Neolithic origins for European paternal lineages." *Public Library of Science Biology* 8(1): 1-9.

Barbieri, Chiara et al. 2012. "Contrasting maternal and paternal histories in the linguistic context of Burkina Faso." *Molecular Biology and Evolution* 29(4): 1213-1223.

Barbieri, Chiara et al. 2016. "Refining the Y chromosome phylogeny with southern African sequences." *Human Genetics* 135: 541-553.

Barker, Graeme et al. 2007. "The 'human revolution' in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo)." *Journal of Human Evolution* 52: 243-261.

Bar-Yosef, Ofer 1987. "Pleisticene connections between Africa and Southwest Asia: an archaeological perspective." *The African Archaeological Review* 5: 29-38.

Bar-Yosef, Ofer 1998. "The Natufian culture in the Levant, threshold to the origins of agriculture." *Evolutionary Anthropology* 6(5): 159-177.

Baskin, Leonid M. 1986. "Differences in the ecology and behavior of reindeer populations in the USSR." *Rangifer* Special Issue No. 1: 333-340.

Baskin, Leonid M. 2003. "River crossings as principal points of human/reindeer relationship in Eurasia." *Rangifer* Special Issue No. 14: 37-40.

Bates, J., C.A. Petriea, and R.N. Singh 2017. "Approaching rice domestication in South Asia: new evidence from Indus settlements in northern India." *Journal of Archaeological Science* 78: 193-201.

Batini, Chiara et al. 2011. "Signatures of the preagricultural peopling processed in Sub-Saharan Africa as revealed by the phylogeography of early Y-chromosome lineages." *Molecular Biology and Evolution* 28(9): 2603-2613.

Battaglia, Vincenza et al. 2009. "Y-chromosomal evidence of the cultural diffusion of agriculture in southeast Europe." *European Journal of Human Genetics* 17: 820-830.

Battaglia, Vincenza et al. 2013. "The first peopling of South America: new evidence from Y-chromosome haplogroup Q." *Public Library of Science One* 8(8): e71390.

Bellwood, Peter 2005. *First farmers: the origins of agricultural societies*. Malden, MA; Oxford, UK; Victoria, Australia: Blackwell Publishing.

Benedict, Paul K. 1987. "Early MY/TB loan relationships." *Linguistics of the Tibeto-Burman Area* 10(2): 12-21.

Bennett, Matthew R. et al. 2021. "Evidence of humans in North America during the Last Glacial Maximum." *Science* 373 (6562): 1528-1531.

Bergstrom, Anders et al. 2016. "Deep roots for Aboriginal Australian Y chromosomes." *Current Biology* 26: 809-813.

Beridze, Tengiz 2019. "The 'Wheat Puzzle' and Kartvelians route to the Caucasus." *Genetic Resources and Crop Evolution* 66: 921-927.

Berniell-Lee, Gemma et al 2009. "Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages." *Molecular Biology and Evolution* 26(7): 1581-1589.

Beyer, Robert M. 2021. "Climatic windows for human migration out of Africa in the past 300,000 years." *Nature Communications* 12:4889.

Binney, Heather et al. 2016. “Vegetation of Eurasia from the last glacial maximum to present: key biogeographic patterns.” *Quaternary Science Reviews* 157: 80-97.

Blench, Roger 2006. *The Niger-Saharan Macrophylum*. Cambridge, UK. Publisher: Author. http://www.rogerblench.info/Language/Nilo-Saharan/NS%20page.htm

Blench, Roger 2010. “Evidence for the Austronesian voyages in the Indian Ocean.” In: *The Global Origins and Development of Seafaring*. Edited by Atholl Anderson et al. Cambridge, UK: McDonald Institute for Archaeological Research, pp. 239-248.

Blench, Roger 2013. “The Prehistory of the Daic- or Kra-Dai-speaking peoples and the hypothesis of an Austronesian Connection.” In: *Unearthing Southeast Asia’s Past: Selected Papers from the 12th International Conference of the European Association of Southeast Asian Archaeologists.* Edited by Marijke J. Klokke and Véronique Degroot. Singapore: NUS Press, pp. 3-15. http://www.jstor.org/stable/j.ctv1qv3nd.6.

Blench, Roger 2014**.** *Linguistic and archaeological evidence for Berber prehistory*.
http://www.rogerblench.info/Language/Afroasiatic/AASOP.htm

Blockley, S.P.E. and R. Pinhasi 2011. “A revised chronology for the adoption of agriculture in the Southern Levant and the role of Late Glacial climatic change.” *Quaternary Science Reviews* 30: 98-108.

Blome, Margaret Whiting et al. “The environmental context for the origins of modern human diversity: A synthesis of regional variability in African climate 150,000-30,000 years ago.” *Journal of Human Evolution* 62(5): 563-592.

Blust, Robert 2013. *The Austronesian Languages*. Revised Edition. Canberra, Australia: Asia-Pacific Linguistics Research School of Pacific and Asian Studies. The Australian National University.

Bocherens, Hervé and Dorothée G. Drucker 2021. “Isotopic insights on ecological interactions between humans and woolly mammoths during the Middle and Upper Palaeolithic in Europe.” In: *Human-Elephant Interactions from Past to Present*. Edited by George E. Konidaris, Ran Barkai, Vangelis Tourloukis and Katerina Harvati. Tübingen University Press, Tübingen, Germany.
http://dx.doi.org/10.15496/publikation-55604

Bolnick, Deborah A. et al. 2004. “Problematic use of Greenberg’s linguistic classification of the Americas in studies of Native American genetic variation*” American Journal of Human Genetics* 75: 519-523.

Bolnick, Deborah A. et al. 2006. “Asymmetric male and female genetic histories among Native Americans from Eastern North America.” *Molecular Biology and Evolution* 23(11): 2161-2174.

Booth, Thomas J. 2019. “A stranger in a strange land: a perspective on archaeological responses to the palaeogenetic revolution from an archaeologist working amongst palaeogeneticists.” *World Archaeology* 51 (4): 586-601.

Bosch, Elena et al. 2001. “High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula.” *American Journal of Human Genetics* 68: 1019-1029.

Bostoen, Koen et al. 2015. “Middle to Late Holocene paleoclimatic change and the early Bantu expansion in the rain forests of western central Africa.” *Current Anthropology* 56(3): 354-384.

Bourgeon, Lauriane, Ariane Burke and Thomas Higham 2017. “Earliest human presence in North America dated to the Last Glacial Maximum: new radiocarbon dates from Bluefish Caves, Canada.” *Public Library of Science One* 12(1): e0169486.

Bowler, James M. et al. 2003. “New ages for human occupation and climatic change at Lake Mungo, Australia.” *Nature* 421: 837-840.

Brackney, Noel C. 2007. *The Origins of Slavonic: Language Contact and Language Change*. Munich: LINCOM Europa.

Brandt, Guido et al. 2015. "Human paleogenetics of Europe. The known knowns and the known unknowns." *Journal of Human Evolution* 79: 73e92.

Brass, Michael 2018. "Early North African cattle domestication and its ecological setting: a reassessment." *Journal of World Prehistory* 31:81–115.

Brown, Wesley M. 1980. "Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis." *Proceedings of the National Academy of Sciences* 77(6): 3605-3609.

Brunelli, Andrea et al. 2017. "Y chromosomal evidence on the origin of northern Thai people." *Public Library of Science One* 12(7): e0181935.

Bruno, Maria C. 2006. "A morphological approach to documenting the domestication of Chenopodium in the Andes." In: *Documenting Domestication: New Genetic and Archaeological Paradigms*. Edited by Melinda A. Zeder et al. Berkeley and Los Angeles: University of California Press, pp. 32-45.

Bučková, Jana et al. 2013. "Multiple and differentiated contributions to the male gene pool of pastoral and farmer populations of the African Sahel." *American Journal of Physical Anthropology* 151: 10-21.

Buvit, Ian and Karisa Terry 2016. "Outside Beringia: why the Northeast Asian Upper Paleolithic record does not support a long standstill model." *Paleo America* 2(4): 281-285.

Cadenas, Alicia M. et al. 2008. "Y-chromosome diversity characterizes the Gulf of Oman." *European Journal of Human Genetics* 16: 374 - 386.

Cai, Xiaoyun et al. 2011. "Human migration through bottlenecks from Southeast Asia into East Asia during the last glacial maximum revealed by Y-chromosomes." *Public Library of Science One* 6(8): e24282.

Campbell, Lyle and Terrence Kaufman 1976. "A linguistic look at the Olmecs." *American Antiquity* 41(1): 80-89.

Campbell, Lyle 1997. *American Indian Languages: the Historical Linguistics of Native America*. New York: Oxford University Press.

Campbell, Lyle and William J. Poser 2008. *Language Classification: History and Method*. Cambridge University Press.

Campbell, Lyle 2011. "The Dené –Yeniseian Connection (Kari and Potter, eds.)." Book review. *International Journal of American Linguistics* 77(3): 445-451.

Cann, Rebecca L., Mark Stoneking and Allan C. Wilson 1987. "Mitochondrial DNA and human evolution." *Nature* 325: 31-36.

Capelli, C. et al. 2006. "Population structure in the Mediterranean Basin: a Y chromosome perspective." *Annals of Human Genetics* 70: 207-225.

Capelli, Cristian et al. 2007. "Y chromosome genetic variation in the Italian peninsula is clinal and supports an admixture model for the Mesolithic-Neolithic encounter." *Molecular Phylogenetics and Evolution* 44: 228-239.

Capredon, Melanie et al. 2013. "Tracing Arab-Islamic inheritance in Madagascar: study of the Y-chromosome and Mitochondrial DNA in the Antemoro." *Public Library of Science On*e 8(11): e80932.

Capriles, Jose M. et al. 2019. "Persistent Early to Middle Holocene tropical foraging in southwestern Amazonia." *Science Advances* 5: eaav5449.

Carrión, J.S. et al. 2019. "The sequence at Carihuela Cave and its potential for research into Neanderthal ecology and the Mousterian in southern Spain." *Quaternary Science Reviews* 217: 194-216.

Casanova, Myriam et al. 1985. "A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance." *Science* 230(4732): 1403-1406.

Cassidy, Lara M. et al. 2016. "Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome." *Proceedings of the National Academy of Sciences of the United States of America* 113(2): 368-373.

Cavalli-Sforza, L.L., Paolo Menozzi, and Alberto Piazza 1994. *The History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.

Cavalli-Sforza, Luigi Luca 2000. *Genes, Peoples and Languages*. Berkeley, CA; Los Angeles, CA: University of California Press.

Chataigner, Christine et al. 2014. "The Neolithic of the Caucasus." *Oxford Handbooks Online*. DOI: 10.1093/oxfordhb/9780199935413.013.13

Chaubey, Gyaneshwer et al. 2011. "Population genetic structure in Indian Austroasiatic speakers: the role of landscape barriers and sex-specific admixture." *Molecular Biology and Evolution* 28(2): 1013-1024.

Chen, Lu et al. 2020. "Identifying and interpreting apparent Neanderthal ancestry in African individuals." *Cell* 180: 677–687.

Chiaroni, Jacques et al. 2008. "Correlation of annual precipitation with human Y-chromosome diversity and the emergence of Neolithic agricultural and pastoral economies in the Fertile Crescent." *Antiquity* 82: 281-289.

Chiaroni, Jacques et al. 2009. "Y-chromosome diversity, human expansion, drift, and cultural evolution." *Proceeding of the National Academy of Sciences of the United States of America* 106(48): 20174-20179.

Chiaroni, Jacques et al. 2010. "The emergence of Y-chromosome haplogroup J1e among the Arabic-speaking populations." *European Journal of Human Genetics* 18: 348-353.

Chritz, Kendra L. et al. 2015. "Environments and trypanosomiasis risks for early herders in the later Holocene of the Lake Victoria basin, Kenya." *Proceedings of the National Academy of Sciences of the United States of America* 112(12): 3674-3679.

Central Intelligence Agency. *World Factbook*. https://www.cia.gov/the-world-factbook/countries/pakistan/ . Accessed October 3, 2021.

Central Intelligence Agency. World Factbook. https://www.cia.gov/the-world-factbook/countries/china/ . Accessed October 3, 2021.

Cinnioğlu, Cengiz et al. 2004. "Excavating Y-chromosome haplotype strata in Anatolia." *Human Genetics* 114: 127-14

Clark, Jeffrey T. and Kevin M. Kelly 1993. "Human genetics, paleoenvironments, and malaria: relationships and implications for the settlement of Oceania." *American Anthropologist* 95(3): 612-630.

Clark, Peter U. et al. 2009. "The Last Glacial Maximum." *Science* 325: 710-714.

Comrie, Bernard 2008. "Linguistic diversity in the Caucasus." *Annual Review of Anthropology* 37: 131-143.

Connor, Simon E. and Eliso V. Kvavadze 2009. "Modelling late Quaternary changes in plant distribution, vegetation and climate using pollen data from Georgia, Caucasus." *Journal of Biogeography* 36(3): 529-545.

Cooper, Alan and Hendrik N. Poinar 2000. "Ancient DNA: do it right or not at all." *Science* 289 (5482): 1139.

Cordaux, Richard et al. 2004. "Independent origins of Indian caste and tribal paternal lineages." *Current Biology* 14: 231-235.

Cox, Francis 2010. "History of the discovery of the malaria parasites and their vectors." *Parasites and Vectors* 3:5.

Crowther, Alison et al. 2017. "Subsistence mosaics, forager-farmer interactions, and the transition to food production in eastern Africa." *Quaternary International* 489: 101-120.

Cruciani, Fulvio et al. 2004. "Phylogenetic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa." *American Journal of Human Genetics* 74: 1014-1022.

Cruciani, Fulvio et al. 2007. "Tracing past human male movements in northern/eastern Africa and western Eurasia: new clues from Y-chromosome haplogroups E-M78 and J-M12." *Molecular Biology and Evolution* 24(6): 1300-1311.

Cruciani, Fulvio et al. 2010. "Human Y chromosome haplogroup R-V88: a paternal genetic record of early mid Holocene trans-Saharan connections and the spread of Chadic languages." *European Journal of Human Genetics* 18: 800-807.

Cui, Yinqiu et al. 2020. "Bioarchaeological perspective on the expansion of Transeurasian languages in Neolithic Amur River basin." *Evolutionary Human Sciences* 2: e15.

Currat, Mathias and Laurent Excoffier 2011. "Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression." *Proceedings of the National Academy of Sciences of the United States of America* 108(37): 15129-15134.

D'Atanasio, Eugenia et al. 2018. "The peopling of the last Green Sahara revealed by high-coverage resequencing of trans-Saharan patrilineages." *Genome Biology* 19: 20.

Darwin, Charles, 1809-1882. *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life*. London: John Murray, 1859.

De Graaf, Tjeerd and Hidetoshi Shiraishi 2013. "Documentation and revitalization of two endangered languages in eastern Asia: Aivkh and Ainu." In: *Sustaining Indigenous Knowledge: Learning Tools and Community Initiatives for Preserving Endangered Languages and Local Cultural Heritage*. Edited by Erich Kasten and Tjeerd de Graaf Fürstenberg/Havel: Kulturstiftung Sibirien, pp. 49-64.

Debnath, Monojit et al. 2011. "Y-chromosome haplogroup diversity in the sub-Himalayan Terai and Duars populations of East India." *Journal of Human Genetics* 56: 765-771.

Decker, Jared E. et al. 2014. "Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle." *Public Library of Science Genetics* 10(3): e1004254.

Delfin, Frederick et al. 2011. "The Y-chromosome landscape of the Philippines: extensive heterogeneity and varying genetics affinities of Negrito and non-Negrito groups." *European Journal of Human Genetics* 19: 224-230.

Delfin, Frederick et al. 2012. "Bridging Near and Remote Oceania: mtDNA and NRY variation in the Solomon Island." *Molecular Biology and Evolution* 29(2): 545-564.

Delfin, Frederick C. 2015. "The population history of the Philippines: a genetic overview." Philippine Studies: *Historical and Ethnographic Viewpoints* 63(4): 449-476.

Demeter, Fabrice et al. 2012. "Anatomically modern human in Southeast Asia (Laos) by 46 ka." *Proceedings of the National Academy of Sciences of the United States of America* 109 (36) 14375-14380

Denham, T.P. et al. 2003. "Origins of agriculture at Kuk Swamp in the highlands of New Guinea." *Science* 301: 189-193.

Dennell, Robin et al. 2020. "A demographic history of Late Pleistocene China." *Quaternary International* 559: 4-13.

Di Cristofaro, Julie et al. 2013. "Afghan Hindu Kush: where Eurasian Sub-Continent gene flows converge." *Public Library of Science One* 8(10): e76748.

Di Cristofaro, Julie et al. 2018. "Prehistoric migrations through the Mediterranean basin shaped Corsican Y chromosome diversity." *Public Library of Science One* 13(8): e0200641.

Di Giacomo, F. et al. 2004. "Y chromosomal haplogroup J as a signature of the post-Neolithic colonization of Europe." *Human Genetics* 115: 357-371.

Dillehay, Tom D. et al. 2015. "New archaeological evidence for an early human presence at Monte Verde, Chile." *Public Library of Science One* 10(11): e0141923.

Dixon, R.M.W. and Alexandra Y. Aikenvald 1999. "Introduction." In: *The Amazonian Languages*. Edited by R.M.W Dixon and Alexandra Y. Aikenvald. Cambridge, UK: Cambridge University Press, pp. 1-21.

Dodykhudoeva, Leyli R. 2003. "The sociolinguistic situation and the state of research of the Ossetic language." First MERCATOR International Symposium on Minority Languages and Research. Abstracts. Wales, United Kingdom, p. 42.

Doi, Tran Tri 2012. "On the relationship between the Austroasiatic and Austronesian languages in Southeast Asia." *Vietnam National University Journal of Social Sciences* 28(5E): 35-39.

Dolukhanov, Pavel M. 2003. "Hunter-gathers of the Last Ice Age in Northern Eurasia." *Before Farming* 2003(2): 1-25.

Dolukhanov, P. 2009. "The Mesolithic of the East European Plain." In: *The East European Plain on the Eve of Agriculture*. Edited by Pavel M. Dolukhanov, Graeme R. Sarson and Anvar M. Shukurov. BAR International Series 1964. Oxford, UK: Archaeopress, pp. 23-34.

Donohue, Mark and Tim Denham 2010. "Farming and language in Island Southeast Asia: reframing Austronesian history." *Current Anthropology* 51(2): 223-256.

Douka, Katerina et al. 2014. "The chronostratigraphy of the Haua Fteah cave (Cyrenaica, northeast Libya)." *Journal of Human Evolution* 66: 39-63.

Drake, Nick A. et al. 2011. "Ancient watercourses and biogeography of the Sahara explain the peopling of the desert." *Proceedings of the National Academy of Sciences of the United States of America* 108(2): 458-462.

Driem, George van 2005. "Tibeto-Burman vs Indo-Chinese: implications for population geneticists, archaeologists and prehistorians." In: *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*." Edited by Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas. London and New York: Routledge, pp. 81-106.

Driem, George van 2011. “Rice and the Austroasiatic and Hmong-Mien homelands.” In: *Dynamics of Human Diversity*. Edited by N. J. Enfield. Canberra: Pacific Linguistics, Australian National University, pp. 361-390.

Driem, George van 2014. “Trans-Himalayan.” In: *Trans-Himalayan Linguistics*. Edited by Nathan Hill and Thomas Owen-Smith. Berlin: Mouton de Gruyter, pp. 11-40.

Dulik, Matthew C. et al. 2012. “Mitochondrial DNA and Y chromosome variation provide evidence for a recent common ancestry between Native Americans and indigenous Altaians.” *American Journal of Human Genetics* 90: 229-246.

Dupuy, Berit Myhre et al. 2006. “Geographical heterogeneity of Y chromosomal lineage in Norway.” *Forensic Science International* 164: 10-19.

Dyke, Arthur S. 2004. “An outline of North American deglaciation with emphasis on central and northern Canada.” *Developments in Quaternary Sciences* 2(B): 373-424.

Egfjord, Anne Friis-Holm et al. 2021. “Genomic steppe ancestry in skeletons from the Neolithic Single Grave Culture in Denmark.” *Public Library of Science One* 16(1): e0244872.

Ehret, Christopher et al. 2004. “The Origins of Afroasiatic.” *Science* 306: 1680-1681.

El-Sibai, Mirvat et al. 2009. “Geographical structure of the Y-Chromosomal genetic landscape of the Levant: a coastal-inland contrast.” *Annals of Human Genetics* 73: 568-581.

Ennafaa, Hajer et al. 2011. “Mitochondrial and Y-chromosome microstructure in Tunisia.” *Journal of Human Genetics* 56: 734-741.

Erlandson, John M., Madonna L. Moss and Matthew Des Lauriers 2008. “Life on the edge: early maritime cultures of the Pacific Coast of North America.” *Quaternary Science Reviews* 27: 2232-2245.

Ethnologue 2014. 17th edition. Edited by Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Ethnologue 2015. 18th edition. Edited by Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Ethnologue 2016. 19th edition. Edited by M. Paul Lewis, Gary F. Simons, and Charles D. Fennig. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Ethnologue 2017. 20th edition. Edited by Simons, Gary F. and Charles D. Fennig (Eds.). Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Ethnologue 2018. 21st edition. Edited by Simons, Gary F. and Charles D. Fennig (Eds.). Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Ethnologue 2019. 22nd edition. Edited by David M. Eberhard, Gary F. Simons, and Charles D. Fennig. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Ethnologue 2021. Edited by Eberhard, David M., Gary F. Simons, and Charles D. Fennig. Twenty-fourth edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

Fadhlaoui-Zid, Karima et al. 2011. “Genetic structure of Tunisian ethnic groups revealed by parental lineages.” *American Journal of Physical Anthropology* 148: 271-280.

Fadhlaoui-Zid, Karima et al. 2013. “Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa.” *Public Library of Science One* 8(11): e80293.

Fehér, T. et al. 2015. "Y-SNP L1034: limited genetic link between Mansi and Hungarian-speaking populations." *Molecular Genetics and Genomics* 290: 377-386.

Fernandes, Daniel M. et al. 2021. "A genetic history of the pre-contact Caribbean." *Nature* 590:103-110.

Fiedel, Stuart 2008. "Sudden deaths: the chronology of terminal Pleistocene megafaunal extinction." In: *American Megafaunal Extinctions at the End of the Pleistocene*. Edited by G. Haynes. Springer, pp. 21-37.

Finocchio, Andrea et al. 2018. "A finely resolved phylogeny of Y chromosome Hg J illuminates the processes of Phoenician and Greek colonizations in the Mediterranean." *Scientific Reports* 8:7465.

Filippo, Cesare de et al. 2011. "Y-chromosomal variation in Sub-Saharan Africa: Insights into the history of Niger-Congo groups." *Molecular Biology and Evolution* 28(3): 1255-1269.

Fischer, Anders 2002. "Food for feasting? An evaluation of explanations for the neolithisation of Denmark and southern Sweden." In: The Neolithisation of Denmark: 150 Years of Debate. Eds. Anders Fischer and Kristian Kristiansen. Sheffield, England: J.R. Collins Publications, pp. 343-393.

Flegontov, Pavel et. al. 2016a. "Na-Dené populations descend from the Paleo-Eskimo migration into America." doi: https://doi.org/10.1101/074476

Flegontov, Pavel et al. 2016b. "Genomic study of the Ket: a Paleo-Eskimo-related ethnic group with significant ancient North Eurasian ancestry." *Scientific Reports* 6:20768.

Flegontov, Pavel et al. 2017. "Paleo-Eskimo genetic legacy across North America." doi: https://doi.org/10.1101/203018

Flegontov, Pavel et al. 2019. "Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America." *Nature* 570 (7760): 236-240.

Florin, S. Anna et al. 2020. "The first Australian plant foods at Madjedbebe, 65,000–53,000 years ago." *Nature Communications* 11: 924.

Forenbaher, Stašo and Preston T. Miracle 2005. "The spread of farming in the Eastern Adriatic." *Antiquity* 79: 514-528.

Fortescue, Michael 2004. "How far west into Asia have Eskimo languages been spoken, and which ones?" *Etudes Inuit Studies* 28(2): 159-183.

Fortescue, Michael 2013. "North America: Eskimo-Aleut linguistic history." In: *The Global Prehistory of Human Migration.* Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 340-345.

Frachetti, Michael D. 2012. "Multiregional emergence of mobile pastoralism and nonuniform institutional complexity across Eurasia." *Current Anthropology* 53(1): 2-38.

Francalacci, Paolo et al. 2015. "Detection of Phylogenetically informative polymorphisms in the entire euchromatic portion of human Y chromosome from a Sardinian sample." *BioMed Central Research Notes* 8:174.

Fregel, Rosa et al. 2018. "Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe." *Proceedings of the National Academy of Sciences* 115(26): 6774-6779.

Friedlaender, Jonathan S. and Serena Tucci 2020. "Human migrations: tales of the Pacific." *Current Biology* 30: R1469–R1499.

Friesen, T. Max 2013. "North America: Paleoeskimo and Inuit archaeology." In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 346-353.

Frison, George C. 1998. “Paleoindian large mammal hunters on the plains of North America.” *Proceedings of the National Academy of Sciences of the United States of America* 95: 14576-14583.

Fromm, Hans 1997. “Germanen in bronzezeitlichen Mittelschweden?” Finnisch-Ugrische Forschungen 54: 127-150.

Frumkin, Amos et al. 2011. “Possible paleohydrologic and paleoclimatic effects on hominin migration and occupation of the Levantine Middle Paleolithic.” *Journal of Human Evolution* 60: 437-451.

Fu, Qiaomei et al. 2014. “Genome sequence of a 45,000-year-old modern human from western Siberia.” *Nature* 514: 445-450.

Fu, Qiaomei et al. 2016. “The genetic history of Ice Age Europe.” *Nature* 534: 200-205.

Fuller, Dorian Q. 2006. “Agricultural origins and frontiers in South Asia: a working synthesis.” *Journal of World Prehistory* 20: 1-86.

Fuller, Dorian Q 2012. “Pathways to Asian civilizations: tracing the origins and spread of rice and rice cultures.” *Rice* 4: 78-92.

Gaffney, Dylan et al. 2021. “Small game hunting in montane rainforests: Specialised capture and broad spectrum foraging in the Late Pleistocene to Holocene New Guinea Highlands.” *Quaternary Science Reviews* 253: 106742.

Garcea, Elena A.A. 2016. “Dispersals Out of Africa and Back to Africa: Modern origins in North Africa.” *Quaternary International* 408 (B): 79-89.

Gavashelishvili, Alexander and David Tarkhnishvili 2016. “Biomes and human distribution during the last Ice Age.” *Global Ecology and Biogeography* 25: 563-574.

Gayden, Tenzin et al. 2007. “The Himalayas as a directional barrier to gene flow.” *American Journal of Human Genetics* 80: 884-894.

Gazi, Nurun Nahar et al. 2013. “Genetic structure of Tibeto-Burman populations of Bangladesh: evaluating the gene flow along the sides of Bay-of-Bengal.” *Public Library of Science One* 8(10): e75064.

Gebremedhin, Gebremeskel Hagos and Abera Asefa Mebrahtu 2020. “Linguistic evolution of Ethiopic language: comparative discussion.” *International Journal of Interdisciplinary Research and Innovations* 8(1): 1-9.

Geppert, Maria et al. 2011. “Hierarchical Y-SNP assay to study the hidden diversity and phylogenetic relationship of native populations in South America.” *Forensic Science International: Genetics* 5: 100-104.

Gilbert, M. Thomas P. et al. 2005. “Assessing ancient DNA studies.” *Trends in Ecology and Evolution* 20(10): 541-544.

Gillispie, Thomas E. 2018. *An Overview of Alaska’s Prehistoric Cultures*. Government publication. Office of History and Archeology Report 173. Anchorage, Alaska: Alaska Department of Natural Resources.

Gimbutas, Marija 1997. “The Fall and Transformation of Old Europe: Recapitulation 1993.” In: *The Kurgan Culture and the Indo-Europeanization of Europe: Selected Articles from 1952-1993*. Eds. Marija Gimbutas; Miriam Robbins Dexter and Karlene Jones-Bley. Washington D.C.: Institute for Study of Man. 351-372.

Glascock, Michael D. et al. 2011. “Obsidian provenance for prehistoric complexes in the Amur River basin (Russian Far East).” *Journal of Archaeological Science* 38: 1832-1841.

Glottolog 4.4. Edited by Hammarström, Harald et al. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.4761960. Available online at http://glottolog.org.

Gokhman, David et al. 2020. "Differential DNA methylation of vocal and facial anatomy genes in modern humans." *Nature Communications* 11:1189.

Golitko, Mark and Lawrence H. Keeley 2007. "Beating ploughshares back into swords: warfare in the Linearbandkeramik." *Antiquity* 81: 332–342.

Gomes, Verónica et al. 2010. "Digging deeper into East African human Y-chromosome lineages. *Human Genetics* 127: 603-613.

Gordon, Bryan 2003. "Rangifer and man: An ancient relationship." *Rangifer*, Special Issue 14: 15-28.

Graf, Kelly E. 2009. "Modern Human Colonization of the Siberian Mammoth Steppe: A View from South-Central Siberia." In: *Sourcebook of Paleolithic Transitions*. Edited by M. Camps and P. Chauhan. Springer Science and Business Media.

Green, Richard E. et al. 2010. "A draft sequence of the Neanderthal genome." *Science* 328: 710-722.

Greenberg, Joseph H. 1987. *Languages in the Americas*. Stanford, CA: Stanford University Press.

Grollemund, Rebecca et al. 2015. "Bantu expansion shows that habitat alters the route and pace of human dispersals." *Proceedings of the National Academy of Sciences of the United States of America* 112(43): 13296–13301.

Grugni, Viola et al. 2012. "Ancient migratory events in the Middle East: new clues from the Y-Chromosome variation of modern Iranians." *Public Library of Science One* 7(7): e41252.

Grugni, Viola et al. 2019. "Analysis of the human Y-chromosome haplogroup Q characterizes ancient population movements in Eurasia and the Americas." *BioMed Central Biology* 17:3.

Guagnin, Maria et al. 2020. "The Holocene humid period in the Nefud Desert: hunters and herders in the Jebel Oraf palaeolake basin, Saudi Arabia." *Journal of Arid Environments* 178: 104146.

Güldemann, Tom and Mark Stoneking 2008. "A historical appraisal of clicks: a linguistic and genetic perspective." *Annual Review of Anthropology* 37: 93-109.

Gusmão, Al. et al. 2008. "A perspective on the history of the Iberian gypsies provided by phylogenetic analysis of Y-chromosome lineages." *Annals of Human Genetics* 72: 215-227.

Haak, Wolfgang et al. 2010. "Ancient DNA from European early Neolithic farmers reveals their Near Eastern affinities." *Public Library of Science Biology* 8 (11).

Haak, Wolfgang et al. 2015. "Massive migration from the steppe was a source for Indo-European languages in Europe." *Nature* 522: 207-211.

Haaland, Randi and Gunnar Haaland 2013. "Early farming societies along the Nile." In: *Oxford Handbook of African Archaeology*. Edited by Peter Mitchell and Paul Lane. Oxford University Press, pp. 542-553.

Haas R. et al. 2017. "Humans permanently occupied the Andean highlands by at least 7 ka." *Royal Society Open Science* 4: 170331.

Haber, Marc et al. 2016. "Chad genetic diversity reveals an African history marked by multiple Holocene Eurasian migrations." *American Journal of Human Genetics* 99: 1316–1324.

Haber, Marc et al. 2019. "A Rare Deep-Rooting D0 African Y-chromosomal haplogroup and its implications for the expansion of modern humans out of Africa." *Genetics* 212: 1421-1428.

Hajdinjak, Mateja et al. 2021. "Initial Upper Palaeolithic humans in Europe had recent Neanderthal ancestry." *Nature* 592: 253-257.

Hallast, Pille et al. 2021. "A Southeast Asian origin for present-day non-African human Y chromosomes." *Human Genetics* 140: 299-307.

Hamilton, Marcus J. and Briggs Buchanan 2010. "Archaeological support for the three-stage expansion of modern humans across northeastern Eurasia and into the Americas." *Public Library of Science One* 5(8): e12472.

Hammer, Michael F. 1994. "A recent insertion of an Alu element on the Y Chromosome is a useful marker for human population studies." *Molecular Biology and Evolution* 11(5): 749-761.

Hammer, Michael F. et al. 2006. "Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y-chromosomes." *Journal of Human Genetics* 51: 47-58.

Hardigan, Michael A. et al. 2017. "Genome diversity of tuber-bearing Solanum uncovers complex evolutionary history and targets of domestication in the cultivated potato." *Proceedings of the National Academy of Sciences of the United States of America* 114 (46): E9999-E10008.

Harris, David R. 1997. "Jeitun and the transition to agriculture in Central Asia." *Archeology International* 1: 28-31.

Harvati, Katerina et al. 2019. "Apidema Cave fossils provide earliest evidence of Homo sapiens in Eurasia." *Nature* 571: 500–504 (2019).

Hassan, Hisham Y. et al. 2008. "Y-chromosome variation among Sudanese: restricted gene flow, concordance with language, geography and history." *American Journal of Physical Anthropology* 137: 316-323.

Haws, Jonathan et al. 2020. "The early Aurignacian dispersal of modern humans into westernmost Eurasia." *Proceedings National Academy of the United States of America* 117(41): 25414-25422.

He, Yungang et al. 2012. "Paleolithic contingent in modern Japanese: estimation and inference using genome-wide data." *Scientific Reports* 2: 355.

He, Jun-Dong et al. 2012. "Patrilineal perspective on the Austronesian diffusion in mainland Southeast Asia." *Public Library of Science One* 7(5): e36437.

Heckenberger, Michael 2013. "Amazonia: archaeology." In: *The Global Prehistory of Human Migration.* Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 392-400.

Heggarty, Paul and David Beresford-Jones 2010. "Agriculture and language dispersals: limitations, refinements, and an Andean exception?" *Current Anthropology* 51(2): 163-193.

Heggarty, Paul and David Beresford-Jones 2013. "Andes: linguistic history." In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 401-409.

Heizer, Robert F. 1944. *Aconite Poison Whaling in Asia and America: an Aleutian Transfer to the New World*. U.S. government document. Anthropological Papers, No. 24. Smithsonian Institution, Bureau of Ethnology, Bulletin 133.

Henn, Brenna M. et al. 2008. "Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa." *Proceedings of the National Academy of Sciences of the United States of America* 105(31): 10693-10698.

Herrera, Kristian J. et al. 2012. "Neolithic patrilineal signals indicate that the Armenian plateau was repopulated by agriculturalists." *European Journal of Human Genetics* 20: 313-320.

Hershkovitz, Israel et al. 2015. “Levantine cranium from Manot Cave (Israel) foreshadows the first European modern humans.” *Nature* 520: 216-219.

Hershkovitz, Israel et al. 2018. “The earliest modern humans outside Africa.” *Science* 359: 456-459.

Hetzron, Robert 2009. Afro-Asiatic languages. In: *The World’s Major Languages. Second Edition*. Bernard Comrie, Ed. Oxon, UK; New York: Routledge. 545-550.

Heupink, Tim H. et al. 2016. “Ancient mtDNA sequences from the First Australians revisited.” *Proceedings of the National Academy of Sciences of the United States of America* 113(25): 6892-6897.

Higham, Charles (2002). “Languages and farming dispersals: Austroasiatic languages and rice cultivation.” In: *Examining the Farming/Language Dispersal Hypothesis*. Edited by Peter Bellwood and Colin Renfrew. Cambridge: McDonald Institute for Archaeological Research, pp. 223-232.

Higham, Charles 2013. “Hunter-gatherers in Southeast Asia: from prehistory to the present.” *Human Biology* 85 (1–3): 21-44.

Hill, A. V. S. et al. 1985. “Melanesians and Polynesians share a unique α-thalassemia mutation.” *American Journal of Human Genetics* 37: 571-580.

Hill, Jane H. 2001. “Proto-Uto-Aztecan: a community of cultivators in Central Mexico?” *American Anthropologist* 103(4): 913-934.

Hirschenfeld, Ludwik and Hanka Hirschenfeld 1919. “Serological differences between the blood of different races.” *The Lancet* 194 (5016): 675-679.

Hoffecker, John F. 2009. “The spread of modern humans in Europe.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (38):16040–16045.

Hofmanova, Zuzana et al. 2016. “Early farmers from across Europe directly descended from Neolithic Aegeans.” *Proceedings of the National Academy of Sciences of the United States of America* 113(25): 6886–6891.

Hofreiter, Michael et al. 2001. “Ancient DNA.” *Nature Reviews: Genetics* 2: 353-359.

Holmes, Charles E. 2011. “The Beringian and transitional periods in Alaska: technology of the East Beringian tradition as viewed from Swan Point.” In: *From the Yenisei to the Yukon: Interpreting Lithic Assemblage Variability in Late Pleistocene/Early Holocene Beringia*. Edited by Ted E. Goebel and Ian Buvit. Center for the Study of the First Americans. College Station, Texas: A&M University Press, pp. 179-191.

Holst Pellekaan, Sheila van 2013. “Genetic evidence for the colonization of Australia.” *Quaternary International* 285: 44-56.

Horborg, Alf 2005. “Ethnogenesis, regional integration, and ecology in prehistoric Amazonia.” *Current Anthology* 16(4): 589-620.

Horsburgh, K. Ann and Mark D. McCoy 2017. “Dispersal, isolation, and interaction in the islands of Polynesia: a critical review of archaeological and genetic evidence.” *Diversity* 9:37.

Housley, R.A. et al. 1997. “Radiocarbon evidence for the late glacial human recolonisation of Northern Europe.” *Proceedings of the Prehistoric Society* 63: 25-54.

Hovhannisyan, Anahit et al. 2014. “Different waves and directions of Neolithic migrations in the Armenian Highland.” *Investigative Genetics* 5:15.

Hu, Kang et al. 2015. “The dichotomy structure of Y chromosome Haplogroup N.” https://arxiv.org/abs/1504.06463

Huang,Yun Zhi et al. 2017. "Whole sequence analysis indicates a recent southern origin of Mongolian Y-chromosome C2c1a1a1-M407." *Molecular Genetics and Genomics* 293(3): 657-663.

Huang, Yun-Zhi et al. 2018. "Dispersals of the Siberian Y-chromosome haplogroup Q in Eurasia." *Molecular Genetics and Genomics* 293:107-117.

Hublin, Jean-Jacques et al. 2017. "New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens." *Nature* 546: 289-292.

Hublin, Jean-Jacques et al. 2020. "Initial Upper Palaeolithic Homo sapiens from Bacho Kiro Cave, Bulgaria." *Nature* 581: 299-302.

Hudjashov, Georgi et al. 2007. "Revealing the prehistoric settlement of Australia by Y chromosome and mtDNA analysis." *Proceeding of the National Academy of Sciences of the United States of America* 104 (21): 8726-8730.

Hudson, Mark J. 2009. "Japanese Beginnings." In: *A Companion to Japanese History*. Edited by William M. Tsutsi. Wiley-Blackwell, pp. 13-29.

Hudson, Mark J. 2013. "Japan: archaeology." In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 224-229.

Hudson, Mark J. 2017a. "The sea and early societies in the Japanese Islands." In: *The Sea in History - The Ancient World.* Edited by C. Buchet, P. Arnaud, and P. De Souza. Boydell and Brewer, pp. 102-113.

Hudson, Mark J. 2017b. "The Ryukyu islands and the northern frontier of prehistoric Austronesian settlement." In: *New Perspectives in Southeast Asian and Pacific Prehistory*. Edited by Philip J. Piper, Hirofumi Matsumura and David Bulbeck. Acton, Australia: Australian National University Press, pp. 189-200.

Hung, Hsiao-chun and Mike T. Carson 2014. "Foragers, fishers and farmers: origins of the Taiwanese Neolithic." *Antiquity* 88: 1115-1131.

Ilumae, Anne-Mai et al. 2016. "Human Y chromosome haplogroup N: a non-trivial time-resolved phylogeography that cuts across language families." *American Journal of Human Genetics* 99: 163-173.

International Society of Genetic Genealogy. http://www.isogg.org.

Issar, Arie S. and Mattanyah Zohar 2007. *Climate Change - Environment and History of the Near East*. 2nd Edition. Berlin and Heidelberg: Springer Verlag.

Janhunen, Juha 2003a. "Para-Mongolic." In: *The Mongolic Languages*. Edited by Juha Janhunen. Routledge: London., pp. 391-402.

Janhunen, Juha 2003b. "Written Mongol." In: *The Mongolic Languages*. Edited by Juha Janhunen. Routledge: London, pp. 30-56.

Janhunen, Juha 2009. "Proto-Uralic—what, where, and when?" *The Quasquicentennial of the Finno-Ugrian Society*. Helsinki: Societe finno-ougrienne. pp. 57-78

Jessen, Catherine A. et al. 2015. "Early Maglemosian culture in the preboreal landscape: archaeology and vegetation from the earliest Mesolithic site in Denmark at Lundby Mose, Sjælland." *Quaternary International* 378: 73-87.

Ji, Xueping et al. 2016. "The oldest Hoabinhian technocomplex in Asia (43.5 ka) at Xiaodong rockshelter, Yunnan Province, southwest China." *Quaternary International* 400: 166-174.

Jobling, Mark A. and Chris Tyler-Smith 2003. "The human Y chromosome: an evolutionary marker comes of age." *Nature Reviews: Genetics* 4: 598-612.

Jochim, Michael et al. 1999. "The Magdalenian colonization of southern Germany." *American Anthropologist* 101:129-142.

Johanson, Lars 2006. "On the roles of Turkic in the Caucasus Area." In: *Linguistic Areas: Convergence in Historical and Typological Perspective*. Edited bu Yaron Matras, April McMahon, and Nigel Vincent. Palgrave Macmillan, pp. 160-181.

Jones, Eppie R. et al. 2015. "Upper Palaeolithic genomes of modern Eurasians." *Nature Communications* 6: 8912.

Jones, Eppie R. et al. 2017. "The Neolithic transition in the Baltic was not driven by admixture with early European farmers." *Current Biology* 27: 576–582.

Jota, Marilza S. et al. 2016. "New native South American Y chromosome lineages." *Journal of Human Genetics* 61(7): 593-603.

Kahlke, Ralf-Dietrich 2015. "The maximum geographic extension of Late Pleistocene Mammuthus primigenius (Proboscidea, Mammalia) and its limiting factors." *Quaternary International* 379: 147-154.

Kane, Daniel 1989. *The Sino-Jurchen Vocabulary of the Bureau of Interpreters*. Bloomington, Indiana: Indiana University Institute for Inner Asian Studies.

Kang, Longli et al. 2012. "Y-chromosome O3 haplogroup diversity in Sino-Tibetan populations reveals two migration routes into the Eastern Himalayas." *Annals of Human Genetics* 76: 92-99.

Karachanak, Sena et al. 2013. "Y-chromosome diversity in modern Bulgarians: new clues about their ancestry." *Public Library of Science One* 8(3): e56779.

Karafet, Tatiana et al. 2008. "New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree." *Genome Research* 18: 830-838.

Karafet, Tatiana M. et al. 2010. "Major east-west division underlies Y-Chromosome stratification across Indonesia." *Molecular Biology and Evolution* 27(8): 1833-1844.

Karafet, Tatiana M. et al. 2015. "Improved phylogenetic resolution and rapid diversification of Y-chromosome haplogroup K-M526 in Southeast Asia." *European Journal of Human Genetics* 23: 369-373.

Karafet, Tatiana et al. 2016. "Coevolution of genes and languages and high levels of population structure among the highland populations of Daghestan." *Journal of Human Genetics* (2016) 61: 181-191.

Karlsson, Andreas O. et al. 2006. "Y-chromosome diversity in Sweden - a long-time perspective." *European Journal of Human Genetics* 14: 963-970.

Karmin, Monika et al. 2015. "A recent bottleneck of Y chromosome diversity coincides with a global change in culture." *Genome Research* 25: 459-466.

Kayser, Manfred et al. 2003. "Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea." *American Journal of Human Genetics* 72: 281-302.

Kayser, Manfred et al. 2006. "Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific." *Molecular Biology and Evolution* 23(11): 2234-2244.

Kayser, Manfred et al. 2008. "The impact of the Austronesian expansion: evidence from mtDNA and Y chromosome diversity in the Admiralty Islands of Melanesia." *Molecular Biology and Evolution* 25(7): 1362-1374.

Keegan, William 2013. "Caribbean Islands: archaeology." In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp 376-383.

Keller, Andreas et al. 2012. “New insights into the Tyrolean Iceman’s origin and phenotype as inferred by whole-genome sequencing.” *Nature Communications* 3: 698.

Kelly, Kevin M. 1990. “GM polymorphisms, linguistics affinities, and natural selection in Melanesia.” *Current Anthropology* 31(2): 201-219.

Kemp, Brian M. et al. 2007. “Genetic analysis of Early Holocene skeletal remains from Alaska and its implications for the settlement of the Americas.” *American Journal of Physical Anthropology* 132: 605-621.

Kemp, Brian M. et al. 2010. “Evaluating the farming/language dispersal hypothesis with genetic variation exhibited by populations in Southwest and Mesoamerica.” *Proceedings of the National Academy of Sciences of the United States of America* 107(15): 6759-6764.

Khlobystin, Leonid P. 2005. *Taymyr: The Archaeology of Northernmost Eurasia*. Edited by William W. Fitzhugh and Vladimir V. Pitulko. Translated from the Russian by Leonid Vishnyatski and Boris Grudinko. Arctic Studies Center, National Museum of Natural History, Smithsonian Institution, Washington, DC.

Khurana, Priyanka et al. 2014. “Y Chromosome haplogroup distribution in Indo-European speaking tribes of Gujarat, Western India.” *Public Library of Science One* 9(3): e90414.

Kılınç, Gulsah Merve et al. 2016. “The demographic development of the first farmers in Anatolia.” *Current Biology* 26(19): 2659-2666.

Kim, Nam-Kil 2009. “Korean.” In: *The World’s Major Languages*. Second edition. Bernard Comrie (Ed.). Oxon, UK; New York: Routledge, pp. 765-779.

Kim, Seung-Og 2015. “Recent developments and debates in Korean prehistoric archaeology.” *Asian Perspectives* 54(1): 11-30.

Kim, Soon-Hee et al. 2011. “High frequencies of Y-chromosome haplogroup O2b-SRY465 lineages in Korea: a genetic perspective on the peopling of Korea.” *Investigative Genetics* 2:10

King, R.J. et al. 2008. “Differential Y-chromosome Anatolian influences on the Greek and Cretan Neolithic.” *Annals of Human Genetics* 72: 205-214.

Kistler, Logan et al. 2018. “Multiproxy evidence highlights a complex evolutionary legacy of maize in South America.” *Science* 362, 1309-1313.

Kivisild, T. et al. 2003. “The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations.” *American Journal of Human Genetics* 72: 313-332.

Kivisild, Toomas 2017. “The study of human Y chromosome variation through ancient DNA.” *Human Genetics* 136(5): 529-546.

Knapp, Michael et al. 2015. “Re-inventing ancient human DNA.” *Investigative Genetics* 6:4.

Knight, Alec et al. 2003. “African Y chromosome and mtDNA divergence provides insight into the history of the click languages.” *Current Biology* 13: 464-473.

Kolpaschikov, Leonid et al. 2015. “The role of harvest, predators, and socio-political environment in the dynamics of the Taimyr wild reindeer herd with some lessons for North America.” *Ecology and Society* 20(1): 9.

Kornfilt, Jaklin 2009. “Turkish and the Turkic languages.” In: *The World’s Major Languages*. Second edition. Bernard Comrie (Ed.). Oxon, UK; New York: Routledge 519-544.

Kosaka, Ryuichi. 2002. "On the affiliation of Miao-Yao and Kadai: can we posit the Miao-Dai family." *Mon-Khmer Studies* 32: 71-100.

Krings, Matthias et al. 1997. "Neandertal DNA sequences and the origin of modern humans." *Cell* 90: 19-30.

Kristensen, Todd J., John W. Ives JW, and Kisha Supernant 2021. "Power, security, and exchange: Impacts of a Late Holocene volcanic eruption in Subarctic North America." *North American Archaeologist* 42(4): 425-472.

Kumar, Vikrant et al. 2007. "Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations." *BioMed Central Evolutionary Biology* 7: 47.

Kuper, Rudolph and Stefan Kroepelin 2006. "Climate-controlled Holocene occupation in the Sahara: motor of Africa's evolution." *Science* 313: 803-807.

Kusuma, Pradiptajati et al. 2015. "Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations." *BioMed Central Genomics* 16: 191.

Kutanan, Wibhu et al. 2019. "Contrasting paternal and maternal genetic histories of Thai and Lao Populations." *Molecular Biology and Evolution* 36(7): 1490-1506.

Kutanan, Wibhu et al. 2020. "Cultural variation impacts paternal and maternal genetic lineages of the Hmong-Mien and Sino-Tibetan groups from Thailand." *European Journal of Human Genetics* 28: 1563-1579.

Kuzmin, Yaroslav V. 2008. "Siberia at the Last Glacial Maximum: environment and archaeology." *Journal of Archaeological Research* 16: 163-221.

Kwon, So Yeun et al. 2015. "Confirmation of Y haplogroup tree topologies with newly suggested Y-SNPs for the C2, O2b and O3a subhaplogroups." *Forensic Science International Genetics* 19: 42-46.

Lacau, Harlette et al. 2012. "Afghanistan from a Y-chromosome perspective." *European Journal of Human Genetics* 20: 1063-1070.

Lahn, Bruce T. et al. 2001. "The human Y chromosome in light of evolution." *Nature Reviews Genetics* 2: 207-216.

Laitinen, Virpi et al. 2002. "Y-chromosomal diversity suggests that Baltic males share common Finno-Ugric-speaking forefathers." *Human Heredity* 53:68-78.

Lander, Faye and Thembi Russell 2018. "The archaeological evidence for the appearance of pastoralism and farming in southern Africa." *Public Library of Science One* 13(6): e0198941.

LaPolla, Randy J. 2001. "The role of migration and language contact in the development of the Sino-Tibetan Language Family." In: *Areal Diffusion and Genetic Inheritance: Case Studies in Language Change*. Edited by R. M. W. Dixon and A. Y. Aikhenvald. Oxford: Oxford University Press, pp. 225-254.

LaPolla, Randy J. 2013. "Eastern Asia: Sino-Tibetan linguistic history." In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 204-208.

LaPolla, Randy J. 2016. "Problems with the arguments for recasting Sino-Tibetan as 'Trans-Himalayan.'" *Linguistics of the Tibeto-Burman Area* 39:(2): 282-297.

Lappalainen, Tuuli et al. 2006. "Regional differences among the Finns: a Y-chromosome perspective." *Gene* 376: 207-215.

Lappalainen, T. et al. 2008. "Migration waves to the Baltic Sea region." *Annals of Human Genetics* 72: 337-348.

Larmuseau, M.H.D. et al 2014. “Increasing phylogenetic resolution still informative for Y chromosomal studies on West-European populations.” *Forensic Science International: Genetics* 9: 179-185.

Lazaridis, Iosif et al. 2016. “Genomic insights into the origin of farming in the ancient Near East.” *Nature* 25: 536 (7617): 419-424.

LeBlanc, Steven A. 2013. “Mesoamerica and the southwestern United States: archaeology.” In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp 369-375.

Lee, Ki-Moon and S. Robert Ramsey 2011. *A History of Korean Language*. Cambridge, UK: Cambridge University Press.

Lewis J.P. et al. 2016. “The shellfish enigma across the Mesolithic-Neolithic transition in southern Scandinavia.” *Quaternary Science Reviews* 151: 315-320.

Li, Dongna et al. 2008. “Paternal genetic structure of Hainan aborigines isolated at the entrance to East Asia.” *Public Library of Science One* 3(5): e2168.

Li, Dong-Na et al. 2013. “Substitution of Hainan indigenous genetic lineage in the Utsat people, exiles of the Champa kingdom.” *Journal of Systematics and Evolution* 51(3): 287-294.

Li, Hui et al. 2007. “Y chromosomes of prehistoric people along the Yangtze River.” *Human Genetics* 122: 383-388.

Li, Mian et al. 2020. “Recent kuru-induced female gene flow disrupted the coevolution of genes and languages in the Papua New Guinea highlands.” *American Journal of Physical Anthropology* 172(1): 87-98.

Linstädter, Jörg 2008. “The Epipalaeolithic-Neolithic-transition in the Mediterranean region of Northwest Africa.” *Quartär* 55: 41-62.

Liu, Wu et al. 2015. “The earliest unequivocally modern humans in southern China.” *Nature* 526: 696–699.

Liu, Bing-Li et al. 2020. “Paternal origin of Tungusic-speaking populations: Insights from the updated phylogenetic tree of Y-chromosome haplogroup C2a-M86.” *American Journal of Human Biology* e23462.

Loikala, Paula. 1977. “The oldest Germanic loanword in Finnish and their contribution to German philology.” *Studi italiani di linguistica teorica ed applicata*. 6 (1-2): 223-250.

Lombard, Marlize 2012. “Thinking through the Middle Stone Age of sub-Saharan Africa.” *Quaternary International* 270:140-155.

López-Parra, A.M. et al. 2009. “In search of the pre- and post-Neolithic genetic substrates in Iberia: Evidence from Y-chromosome in Pyrenean populations.” *Annals of Human Genetics* 73: 42-53.

Loy, Dorothy E. et al. 2018. “Evolutionary history of human Plasmodium vivax revealed by genome-wide analyses of related ape parasites.” *Proceedings of the National Academy of Sciences of the United States of America* 115(36): E8450–E8459.

Luis, J.R. et al. 2004. “The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations.” *American Journal of Human Genetics* 74: 532-544.

Macholdt, Enrico et al. 2020. “The paternal and maternal genetic history of Vietnamese populations.” *European Journal of Human Genetics* 28: 636–645.

Mailhammer, Robert 2007. *The Germanic Strong Verbs: Foundations and Development of a New System*. Berlin; New York: Mouton de Gruyter.

Maley, Jean et al. 2018. "Late Holocene forest contraction and fragmentation in central Africa." *Quaternary Research* 89: 43–59.

Malhi, Ripan Singh et al. 2008. "Distribution of Y chromosomes among native north Americans: a study of Athabaskan population history." *American Journal of Physical Anthropology* 137: 412-424.

Mallory, J., A. Dybob, and O. Balanovsky 2019. "The Impact of Genetics Research on Archaeology and Linguistics in Eurasia." *Russian Journal of Genetics* 55(12): 1472-1487.

Mann, Daniel H. et al. 2015. "Life and extinction of megafauna in the ice-age Arctic." *Proceedings of the National Academy of Sciences of the United States* 112(46): 14301-14306.

Marcheco-Teruel, Beatriz et al. 2014. "Cuba: exploring the history of admixture and the genetic basis of pigmentation using autosomal and uniparental markers." *Public Library of Science Genetics* 10(7): e1004488.

Martinón-Torres, María et al. 2021. "Earliest known human burial in Africa." *Nature* 593: 95-100.

Martins, Haidé et al. 2015. "Radiocarbon Dating the Beginning of the Neolithic in Iberia: New Results, New Problems." *Journal of Mediterranean Archaeology* 28(1): 105-131.

Martirosyan, Hrach 2013. "The place of Armenian in the Indo-European language family: the relationship with Greek and Indo-Iranian." *Journal of Language Relationship* 10: 85-138.

Matson, R. G. and M. P. R. Magne 2013. "North America: Na Dené /Athapaskan archeology and linguistics." In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 333-339.

McColl, Hugh. et al. 2018. "The prehistoric peopling of Southeast Asia." *Science* 361 (6397):88-92.

McEvoy, Brian et al. 2004. "The longue durée of genetic ancestry: multiple genetic marker systems and Celtic origins on the Atlantic façade of Europe." *American Journal of Human Genetics* 75: 693-702.

Mellars, Paul 2006. "Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia." *Science* 313: 796-800.

Mendez, Fernando L. et al. 2011. "Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa." *Human Biology* 83(1): 39-53.

Mendez, Fernando L. et al. 2013. "An African American paternal lineage adds an extremely ancient root to the human Y chromosome phylogenetic tree." *American Society of Human Genetics* 92: 454-459.

Mendez, Fernando L. et al. 2016. "The divergence of Neandertal and modern human Y chromosomes." *American Journal of Human Genetics* 98: 728-734.

Mengoni Gonalons, Guillermo L. and Hugo D. Yaco Baccio 2006. "The domestication of South American camelids: a view from the South-Central Andes." In: *Documenting Domestication: New Genetic and Archaeological Paradigms*. Edited by Melinda A.Zeder et al. Berkeley and Los Angeles: University of California Press, pp. 228-244.

Merrill, William L. 2009. "The diffusion of maize to the southwestern United States and its impact." *Proceedings of the National Academy of Sciences of the United States* 106(50): 21019-21029.

Meyer, Christian et al. 2015. "The massacre mass grave of Schöneck-Kilianstädten reveals new insights into collective violence in Early Neolithic Central Europe." *Proceedings of the National Academy of Sciences of the United States of America* 112(36): 11217–11222.

Meyer, Christian et al. 2018. "Early Neolithic executions indicated by clustered cranial trauma in the mass grave of Halberstadt." *Nature Communications* 9:2472.

Militarev, Alexander 2002. "The prehistory of a dispersal: the Proto-Afrasian (Afroasiatic) farming lexicon." In: *Examining the Farming/Language Dispersal Hypothesis*. Edited by Peter Bellwood and Colin Renfrew. Cambridge: McDonald Institute for Archaeological Research, pp. 135-150.

Mirov, N. T. 1945. "Notes on the domestication of reindeer." *American Anthropologist* 47: 393-408.

Misra, Virendra Nath 2001. "Prehistoric human colonization of India." *Journal of Biosciences* 26(4):491-531.

Mitchell, Peter 2010. "Genetics and southern African prehistory: an archaeological view." *Journal of Anthropological Sciences* 88: 73-92.

Mona, Stefano et al. 2007. "Patterns of Y-chromosome diversity intersect with the Trans-New Guinea hypothesis." *Molecular Biology and Evolution* 24(11): 2546-2555.

Mona, Stefano et al. 2009. "Genetic admixture history of Eastern Indonesia as revealed by Y-chromosome and mitochondrial DNA analysis." *Molecular Biology and Evolution* 26(8): 1865-1877.

Mondal, Mayukh et al. 2017. "Y-chromosomal sequences of diverse Indian populations and the ancestry of the Andamanese." *Human Genetics* 136(5):499-510.

Montano, Valeria et al. 2011. "The Bantu expansion revisited: a new analysis of Y chromosome variation in Central Western Africa." *Molecular Ecology* 20: 2693-2708.

Morein, Eugene 2008. "Evidence for declines in human population densities during the early Upper Paleolithic in western Europe." *Proceedings of the National Academy of Sciences of the United States of America* 105(1): 48-53.

Moreno-Mayar, J. Víctor et al. 2018. "Early human dispersals within the Americas." *Science* 362 (6419): eaav2621.

Moss, Madonna L. and Jon M. Erlandson 1995. "Reflections on North American Pacific coast prehistory." *Journal of World Prehistory* 9(1): 1-45.

Müller, Ulrich C. et al. 2011. "The role of climate in the spread of modern humans into Europe." *Quaternary Science Reviews* 30: 273-279.

Mulligan, Connie J. and Emoke J.E. Szathmary 2017. "The peopling of Americas and the origin of the Beringian occupation model." *American Journal of Physical Anthropology* 162: 403-408.

Myres, Natalie M. et al. 2011. "A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe." *European Journal of Human Genetics* 19: 95-101.

Nagle, Nano et al. 2016a. "Antiquity and diversity of aboriginal Australian Y-chromosomes." *American Journal of Physical Anthropology* 159: 367-381.

Nagle, Nano et al. 2016b. "Mitochondrial DNA diversity of present-day aboriginal Australians and implications for human evolution in Oceania." *Journal of Human Genetics* 62: 343-353.

Naidoo, Thijessen et al. 2020. "Y-Chromosome Variation in Southern African Khoe-San Populations Based on Whole-Genome Sequences." *Genome Biology and Evolution* 12(7):1031–1039.

Naitoh, Sae et al. 2013. "Assignment of Y-chromosomal SNPs found in Japanese population to Y-chromosomal haplogroup tree." *Journal of Human Genetics* 58: 195-201.

Narasimhan, Vagheesh M. et al. 2019. "The formation of human populations in South and Central Asia." *Science* 365: eaat7487.

Nasidze, Ivan et al. 2003. "Testing hypotheses of language replacement in the Caucasus: evidence from the Y-Chromosome." *Human Genetics* 112: 255-261.

Nasidze, I. et al. 2004a. "Mitochondrial and Y-chromosome variation in the Caucasus." Annals of Human Genetics 68: 205-221.

Neves, Walter Alves, Diogo Meyer, and Hector Mario Pucciarella 1996. "Early skeletal remains and the peopling of the Americas." *Revista de Anthropologia* 39(2): 121-139.

Nikolskiy, Pavel and Vladmir Pitulko 2013. "Evidence from the Yana Palaeolithic site, Arctic Siberia, yields clues to the riddle of mammoth hunting." *Journal of Archaeological Science* 40: 4189-4197.

Ning, Chao et al. 2016. "Refined phylogenetic structure of an abundant East Asian Y-chromosomal haplogroup O*-M134." *European Journal of Human Genetics* 24: 307-309.

Nishiaki, Yoshihiro et al. 2015. "Chronological contexts of the earliest pottery Neolithic in the South Caucasus: Radiocarbon dates for Götepe and Haci Elamxanli Tepe, Azerbaijan." *American Journal of Archaeology* 119(3): 279-294.

Nogués-Bravo David et al. 2008. "Climate change, humans, and the extinction of the woolly mammoth." *Public Library of Science Biology* 6(4): e79.

Norquest, Peter K. 2007. *A Phonological Reconstruction of Proto-Hlai*. Dissertation. University of Arizona, 2007.

Olalde, Inigo et al 2018. "The beaker phenomenon and the genomic transformation of Northwest Europe." *Nature* 555: 190-196.

Olsen, Kenneth M. and Barbara A. Schaal 1999. "Evidence on the origins of cassava: Manihot esculenta." *Proceedings of the National Academy of Sciences of the United States of America* 96: 5586-5591.

Oppenheimer, Stephen 2012. "Out-of-Africa, the peopling of continents and islands: tracing uniparental gene trees across the map." *Philosophical Transactions of the Royal Society B: Biological Sciences* 367: 770-784.

Orlova, L.A. et al. 2001. "Chronology and environment of woolly mammoth (Mammuthus primigenius Blumenbach) extinction in northern Asia." *The World of Elephants - International Congress*, Rome 2001

Ostapirat, Weera 2018. "Macrophyletic trees of East Asian languages re-examined." *Senri Ethnological Studies* 98: 107-121.

Ottoni, Claudio et al. 2011. "Deep into the roots of the Libyan Tuareg: a genetic survey of their paternal heritage." *American Journal of Physical Anthropology* 145: 118-124.

Oven, Mannis van and Manfred Kayser 2008. "Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation." *Human Mutation* 30(2): E386-E394.

Oven, Mannis van et al. 2014. "Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome." *Human Mutation* 35:187-191.

Pääbo, Svante 1985. "Molecular cloning of ancient Egyptian mummy DNA." *Nature* 314: 644-645.

Pääbo, Svante et al. 2004. "Genetic analyses from ancient DNA." *Annual Review of Genetics* 38: 645-679.

Pakendorf, Brigitte et al. 2006. "Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of the Yakuts." *Human Genetics* 120: 334-353.

Pamjav, Horolma et al. 2011. "Genetic structure of the parental lineage of the Roma people." *American Journal of Physical Anthropology* 145: 21-29.

Pamjav, Horolma et al. 2012. "New Y-chromosome binary markers improve phylogenetic resolution within haplogroup R1a1." *American Journal of Physical Anthropology* 149: 611-615.

Park, Myung Jin et al. 2012. "Understanding the Y chromosome variation in Korea—relevance of combined haplogroup and haplotype analyses." *International Journal of Legal Medicine* 126: 589–599.

Parton, Ash et al. 2015. "Orbital-scale climate variability in Arabia as a potential motor for human dispersals." *Quaternary International* 382 82-97.

Pavlov, B.M. et al. 1994. "Population dynamics of the Taimyr reindeer population." *Rangifer* Special Issue No. 9.

Pawley, Andrew 2005. "The chequered career of the Trans New Guinea hypothesis: recent research and its implications." In: *Papuan Pasts: Cultural, Linguistic and Biological Histories of Papuan-Speaking Peoples*. Edited by Andrew Pawley et al. Canberra, Australia: Pacific Linguistics, pp. 67-107.

Pedersen, Holger 1967. *The Discovery of Language: Linguistic Science in the Nineteenth Century.* Translated by John Webster Spargo. Bloomington: Indiana University Press.

Pedro, Nicole et al. 2020. "Papuan mitochondrial genomes and the settlement of Sahul." *Journal of Human Genetics* 65: 875–888.

Pereira, Luísa et al. 2010. "Linking the sub-Saharan and West Eurasian gene pools: maternal and paternal heritage of the Tuareg nomads from the African Sahel." *European Journal of Human Genetics* 18: 915-923.

Petrejčíková, Eva et al. 2009. "Y-haplogroup frequencies in the Slovak Romany population." *Anthropological Science* 117(2): 89-94.

Petrosyan, Armen 2007. "The problem of identification of the proto-Armenians: a critical review." *Journal of the Society for Armenian Studies* 16: 25-66.

Pfeifer, Sebastian J. et al. 2019. "Mammoth ivory was the most suitable osseous raw material for the production of Late Pleistocene big game projectile points." *Scientific Reports* 9:2303.

Piličiauskas, Gytis et al. 2019. "Comb ware culture in Lithuania: new evidence from Šventoji 43." *Lietuvos Archeologija* 45: 67-103

Piličiauskas, Gytis et al. 2020. "Fishers of the Corded Ware Culture in the Eastern Baltic." *Acta Archaeologica* 91: 95-120.

Pinotti, Thomaz et al. 2019. "Y chromosome sequences reveal a short Beringian standstill, rapid expansion, and early population structure of Native American founders." Current Biology 29: 149-157.

Pitulko, V. V. and P. A. Nikolskiy 2012. "The extinction of the woolly mammoth and the archaeological record in Northeastern Asia." *World Archaeology* 44(1): 21-42.

Pitulko, V.V. et al. 2004. "The Yana RHS site: humans in the Arctic before the Last Glacial Maximum." *Science* 303: 52-56.

Pitulko, Vladimir V. et al. 2016. "Early presence in the Arctic: evidence from 45,000-year-old mammoth remains." *Science* 351(6270): 260-263.

Platt, Daniel E. et al. 2017. "Mapping post-glacial expansions: the peopling of Southwest Asia." *Scientific Reports* 7: 40338.

Platt, Daniel E. et al. 2021. "Autosomal genetics and Y-chromosome haplogroup L1b-M317 reveal Mount Lebanon Maronites as a persistently non-emigrating population." *European Journal of Human Genetics* 29: 581-592.

Poetsch, Micaela et al. 2013. "Determination of population origin: a comparison of autosomal SNPs, Y-chromosomal and mtDNA haplogroups using a Malagasy population as example." *European Journal of Human Genetics* 21: 1423-1428.

Pokharia, Anil K. et al. "Altered cropping pattern and cultural continuation with declined prosperity following abrupt and extreme arid event at ~4,200 yrs BP: Evidence from an Indus archaeological site Khirsara, Gujarat, western India." *Public Library of Science One* 12(10): e0185684.

Pope, Kevin O. and John E. Terrell 2008. "Environmental setting of human migrations in the circum-Pacific region." *Journal of Biogeography* 35: 1-21.

Potter, Ben A. et al. 2014. "New insights into Eastern Beringian mortuary behavior: a terminal Pleistocene double infant burial at Upward Sun River." *Proceedings of the National Academy of Sciences of the United States of America* 111(48): 17060-17065.

Potter, Ben A. et al. 2018. "Current evidence allows multiple models for the peopling of the Americas." *Science Advances* 4: eaat5473.

Poznik, G. David et al. 2016. "Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences." *Nature Genetics* 48(6): 593-600.

Pringle, Heather 2006. *The Master Plan: Himmler's Scholars and the Holocaust.* New York: Hyperion.

Prüfer, Kay et al. 2021. "A genome sequence from a modern human skull over 45,000 years old from Zlatý kůň in Czechia." *Nature Ecology and Evolution*. 5: 820–825.

Puzachenko, A.Yu. and A.K. Markova 2019. "Evolution of mammal species composition and species richness during the Late Pleistocene - Holocene transition in Europe: a general view at the regional scale." *Quaternary International* in press.

Qi, Xuebin et al. 2013. "Genetic evidence of Paleolithic colonization and Neolithic expansion of modern humans on the Tibetan Plateau." *Molecular Biology and Evolution* 30(8): 1761-1778.

Qiu, Qiang et al. 2015. "Yak whole-genome resequencing reveals domestication signatures and prehistoric population expansions." *Nature Communications* 6: 10283.

Quintana-Murci, Luís et al. 2008. "Maternal traces of deep common ancestry and asymmetric gene flow between Pygmy hunter-gatherers and Bantu-speaking farmers." *Proceedings of the National Academy of Sciences of the United States of America* 105(5): 1596-1601.

Raghavan, Maanasa et al. 2014. "Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans." *Nature* 505: 87-91.

Rai, Niraj et al. 2012. "The phylogeography of Y-chromosome haplogroup H1a1a-M82 reveals the likely Indian origin of the European Romani populations." *Public Library of Science One* 7(11): e 48477.

Rampino, Michael R. and Stanley H. Ambrose 2000. "Volcanic winter in the Volcanic winter in the Garden of Eden: The Toba supereruption and the late Pleistocene human population crash." *Geological Society of America Special Papers* 345: 71-82.

Rasmussen, Morten et al. 2010. "Ancient human genome sequence of an extinct Palaeo-Eskimo." *Nature* 463: 757-762.

Rasmussen, Morten et al. 2014. "The genome of a Late Pleistocene human from a Clovis burial site in western Montana." *Nature* 506: 225-229.

Rasmussen, Morten et al. 2015. "The ancestry and affiliations of Kennewick Man." *Nature* 523: 455-458.

Ratliff, Martha 2010. *Hmong-Mien Language History*. Canberra, Australia: Pacific Linguistics.

Reardon, Sara 2017. "Navajo Nation reconsidered ban on genetic research." *Nature* 550: 165-166.

Redd, Alan J. et al. 2002. "Gene flow from the Indian Subcontinent to Australia: evidence from the Y Chromosome." *Current Biology* 12: 673-677.

Regueiro, Maria et al. 2012. "*High levels of Paleolithic Y-chromosome lineages characterize Serbia*." *Gene* 498: 59-67.

Regueiro, Maria et al. 2013. "On the origins, rapid expansion and genetic diversity of Native Americans from hunting-gatherers to agriculturalists." *American Journal of Physical Anthropology* 150: 333-348.

Reich, David et al. 2011. "Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania." *American Journal of Human Genetics* 89: 516–528.

Reich, David 2020. Dataset Release V42.4 (March 1, 2020). https://reich.hms.harvard.edu/

Renfrew, Colin 1987. *Archaeology and Language: the Puzzle of Indo-European Origins*. Cambridge; New York: Cambridge University Press.

Renfrew, Colin 1989. "The origins of Indo-European languages." *Scientific American* 261(4): 106-115.

Ricankova, Vera Pavelkova et al. 2014. "Ecological structure of recent and Last Glacial mammalian faunas in Northern Eurasia: the case of Altai-Sayan refugium." *Public Library of Science One* 9(1): e85056.

Robbeets, Martine 2008. The historical comparison of Japanese, Korean and the Trans-Eurasian languages. *Rivista degli studi orientali* 81:(1-4): 261-287.

Robbeets, Martine 2017a. "Proto-Trans-Eurasian: Where and When?" *Man in India* 97(1): 19-46.

Robbeets, Martine 2017b. "Austronesian influence and Transeurasian ancestry in Japanese." *Language Dynamics and Change* 7: 210-251.

Roberts, Patrick and Brian A. Stewart 2018. "Defining the 'generalist specialist' niche for Pleistocene Homo sapiens." *Nature Human Behaviour* 2: 542-550.

Rocha, Jorge and Anne-Maria Fehn 2016. "Genetics and demographic history of the Bantu." In: Wiley eLS Online Library. http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0022892/full

Røed, Knut H. et al. 2008. "Genetic analyses reveal independent domestication origins of Eurasian reindeer." *Proceedings of the Royal Society B* 275: 1849-1855.

Roewer, Lutz et al. 2013. "Continent-wide decoupling of Y-chromosomal genetic variation from language and geography in Native South Americans." *Public Library of Science Genetics* 9(4): e1003460.

Rohrlach, Adam B.et al. 2021. "Using Y-chromosome capture enrichment to resolve haplogroup H2 shows new evidence for a two-path Neolithic expansion to Western Europe." *Scientific Reports* 11: 15005.

Roosevelt, A.C. et al. 1996. "Paleo-Indian cave dwellers in the Amazon: the peopling of the Americas." *Science* 272: (5260): 373-384.

Rootsi, Siri et al. 2004. "Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe." *American Journal of Human Genetics* 75: 128-137.

Rootsi, Siiri et al. 2007. "A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe." *European Journal of Human Genetics* 15: 204-211.

Rootsi, Siiri et al. 2012. "Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus." *European Journal of Human Genetics* 20: 1275-1282.

Rosa, Alexandra et al. 2007. "Y-chromosome diversity in the population of Guinea-Bissau: a multiethnic perspective." *BioMed Central Evolutionary Biology* 7: 124.

Roullier, Caroline et al. 2013. "Disentangling the origins of cultivated sweet potato." *Public Library of Science One* 8(5): e62707.

Rowold, Daine J. et al. 2016. "On the Bantu expansion." *Gene* 593: 48-57.

Rowald, Diane J. et al. 2020. "The Y-chromosome of the Soliga, an ancient forest-dwelling tribe of South India." Gene 763 (Supplement):100026.

Ruhlen, Merritt 1998. "The origin of Na-Dené." *Proceedings of the National Academy of Sciences of the United States of America* 95: 13994-13996.

Ryan, William B.F. et al 2003. "A catastrophic flooding of the Black Sea." *Annual Review of Earth and Planetary Sciences*. 31: 525-554.

Ryan, William B.F. et al. 1997. "An abrupt drowning of the Black Sea shelf." *Marine Geology* 138: 119-126.

Rybin, Evgeny et al. 2016. "The impact of the LGM on the development of the Upper Paleolithic in Mongolia." *Quaternary International* 425: 69-87.

Sagart, Laurent 2004. "The higher phylogeny of Austronesian and the position of Tai-Kadai." *Oceanic Linguistics* 43(2): 411-444.

Sahoo, Sanghamitra et al. 2006. "A prehistory of Indian Y-chromosomes: evaluating demic diffusion scenarios." *Proceedings of the National Academy of Sciences of the United States of America* 103(3): 843-848.

Salminen, Tapani 2007. "The Rise of the Finno-Ugric language family." In: *Early Contacts between Uralic and Indo-European: Linguistic and Archaeological Considerations*. Second edition. Edited by Christian Carpelan et al. Vammala, Finland: The Finno-Ugrian Society and Authors, pp.385-396.

Sanchez, J.J. et al. 2004. "Y chromosome SNP haplogroups in Danes, Greenlanders and Somalis." *International Congress Series* 1261: 347-349.

Sankararaman, Sriram et al. 2012. "Date of interbreeding between Neandertals and modern humans." *Public Library of Science Genetics* 8(10): e1002947.

Sankararaman, Sriram et al. 2014. "The genomic landscape of Neanderthal ancestry in present-day humans." *Nature* 507 354–357 (2014).

Sankararaman, Sriram 2020. "Methods for detecting introgressed archaic sequences." *Current Opinion in Genetics and Development* 62: 85–90.

Sarac, Jelena et al. 2016. "Genetic heritage of Croatians in the Southeastern European gene pool - Y chromosome analysis of the Croatian continental and island population." *American Journal of Human Biology* 28:837–845.

Sato, Youichi et al. 2014. "Overview of genetic variation in the Y chromosome of modern Japanese males." *Anthropological Science* 122(3): 131-136.

Savelle, James M. and Nobuhiro Kishigami 2013. "Anthropological research on whaling: prehistoric, historic and current contexts." *Senri Ethnological Studies* 84: 1-48.

Schapper, Antoinette 2017. "Farming and the Trans-New Guinea family: a consideration." In: *Language Dispersal beyond Farming*. Edited by Martine Robbeets and Alexander Savelyev. John Benjamins Publishing Company, pp. 155-181.

Schirmer, Alfred and Walther Mitzka 1969. *Deutsche Wortkunde* 6th edition. Berlin: Walter de Gruyter.

Schleicher, August 1863. *Die Darwinsche Theories und die Sprachwissenschaft: Offenes Sendsschreiben an Herrn Dr. Ernst Häcket*. Weimar: Hermann Böhlau.

Schurr, Theodore G. et al. 2012. "Clan, language, and migration history has shaped genetic diversity in Haida and Tlingit populations from Southeast Alaska." *American Journal of Physical Anthropology* 148: 422-435.

Scozzari, Rosaria et al. 2012. "Molecular dissection of the basal clades in the human Y chromosome phylogenetic tree." *Public Library of Science One* 7(11): e49170.

Semino, Ornella et al. 1996. "A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers." *American Journal of Human Genetics* 59: 964-968.

Semino, Ornella et al 2004. "Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the Neolithization of Europe and later migratory events in the Mediterranean area." *American Journal of Human Genetics* 74: 1023-1034.

Sengupta, Sanghamitra et al. 2006. "Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveals minor genetic influence of central Asian pastoralists." *American Journal of Human Genetics* 78: 201-221.

Serdyuk, Natalia 2005. "The history of mammalian communities and paleography of Altai mountains in the Paleolithic." *Paleontological Journal* 39 (Suppl. 6): S645–S821.

Shang, Hong et al. 2007. "An early modern human from Tianyuan Cave, Zhoukoudian, China." *Proceedings of the National Academy of Sciences of the United States of America* 104(16): 6573-6578.

Shi, Hong et al. 2013. "Genetic evidence of an East Asian origin and Paleolithic northward migration of Y-chromosome haplogroup N." *Public Library of Science One* 8(6): e66102.

Shibatani, Masayoshi 2009. "Japanese." In: *The World's Major Languages*. Second edition. Edited by Bernard Comrie. Oxon, UK; New York: Routledge, pp. 741-763.

Shimoji, Michinori 2010. "Ryukyuan languages: an introduction." In: An Introduction to Ryukyuan Languages. Edited by Michinori Shimoji and Thomas Pellard. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, pp. 1-13.

Sidwell, Paul 2010. "The Austroasiatic central riverine hypothesis." *Journal of Language Relationship* 4: 117-134.

Sidwell, Paul 2013. "Southeast Asian mainland: linguistic history." In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 259-268.

Siiräinen, Ari 2003. "The Stone and Bronze Ages." In: *The Cambridge History of Scandinavia*, Vol. 1. Knut Helle (Ed). Cambridge University Press, pp. 43-59.

Sikora, Martin et al. 2017. "Ancient genomes show social and reproductive behavior of early Upper Paleolithic foragers." *Science* 3:358: 659-662.

Sikora, Martin et al. 2018. "The population history of northeastern Siberia since the Pleistocene." *Nature* 570: 182-188.

Silva, Fabio et al. 2015. "Modelling the geographical origin of rice cultivation in Asia using the rice archaeological database." *Public Library of Science One* 10(9): e0137024.

Singh, Sakshi et al. 2016. "Dissecting the influence of Neolithic demic diffusion on Indian Y-chromosome pool through J2-M172 haplogroup." *Scientific Reports* 6: 19157.

Singh, Prajjval Pratap et al. 2021. "Dissecting the paternal founders of Mundari (Austroasiatic) speakers associated with the language dispersal in South Asia." *European Journal of Human Genetics* 29: 528–532.

Skoglund, Pontus et al. 2015. "Genetic evidence for two founding populations of the Americas." *Nature* 525: 104-110.

Slatkin, Montgomery and Fernando Racimo 2016. "Ancient DNA and human history." *Proceedings of the National Academy of Sciences of the United States of America* 113(23): 6380-6387.

Smalley, John and Michael Blake 2003. "Sweet beginnings: stalk sugar and the domestication of maize." *Current Anthropology* 44(5): 675-703.

Smith, Andrew B. 1992. "Origins and Spread of Pastoralism in Africa." *Annual Review of Anthropology* 21: 125-141.

Snow, Dean 2013. "Eastern North America: archeology and linguistics." In: *The Global Prehistory of Human Migration*. Edited by Peter Bellwood. West Sussex, UK: John Wiley and Sons, pp. 354-361.

Solé-Morata, Neus et al. 2017. "Analysis of the R1b-DF27 haplogroup shows that a large fraction of Iberian Y-chromosome lineages originated recently in situ." *Scientific Reports* 7: 7341.

Sommer, Robert S. et al. 2014. "Range dynamics of the reindeer in Europe during the last 25,000 years." *Journal of Biogeography* 41: 298-306.

Spengler, Robert et al. 2014. "Early agriculture and crop transmission among Bronze Age mobile pastoralists of Central Eurasia." *Proceedings of the Royal Society B* 281: 20133382.

Spooner, David M . et al. 2005. "A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping." *Proceedings of the National Academy of Sciences of the United States of America* 102 (41): 14694–14699.

Sporrong, Ulf 2003. "The Scandinavian landscape and its resources." In: *The Cambridge History of Scandinavia*, vol. 1. Edited by Knut Helle. *Cambridge University Press*, pp. 15-42.

St. Clair, Michael R. 2012. *Germanic Origins from the Perspective of the Y-chromosome.*
Ph.D. dissertation. http://oskicat.berkeley.edu/record=b20247897~S1

St. Clair, Michael R. 2014. "Population genetics and the humanities: Crimean Goths and contemporary European Romani." *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis*. 19(1): 135-142.

St. Clair, Michael R. 2017. "The co-evolution of agriculture, language and human genetic mutations." *Interdisciplinary Journal for Germanic Linguistics and Semiotic Analysis* 22(2): 229-239.

Stevens, Chris et al. 2016. "Between China and South Asia: A Middle Asian corridor of crop dispersal and agricultural innovation in the Bronze Age." *The Holocene* 26(10): 1541-1555.

Stevens, Chris J. and Dorian Q. Fuller 2017. "The spread of agriculture in eastern Asia: archaeological bases for hypothetical farmer/language dispersals." *Language Dynamics and Change* 7: 152-186.

Stevens, Rhiannon et al. 2009. "Palaeoenvironmental and chronological investigations of the Magdalenian sites of Goyet Cave and Trou de Chaleux (Belgium), via stable isotope and radiocarbon analyses of horse skeletal remains." *Journal of Archaeological Science* 36: 653–662.

Stewart, J. R. and C. B. Stringer 2012. "Human evolution out of Africa: the role of refugia and climate change." *Science* 1317-1321.

Stoneking, Mark et al. 1990. "Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124: 717-733.

Stoneking, Mark and Frederick Delfin 2010. "The human genetic history of East Asia: weaving a complex tapestry." *Current Biology* 20: R188-R193.

Straus, Lawrence Guy, David J. Meltzer and Ted Goebel 2005. "Ice Age Atlantis? Exploring the Solutrean-Clovis 'connection.'" *World Archaeology* 37(4): 507-532.

Suihkonen, Pirkko (2002). The Uralic languages. *Fennia* 180(1-2): 165–176.

Summerhayes, Glenn R. and Atholl Anderson 2009. "An Austronesian presence in southern Japan: early occupation in the Yaeyama Islands." *Bulletin of the Indo-Pacific Prehistory Association* 29: 76-91.

Sun, Na et al. 2019. "Phylogeography of Y-chromosome haplogroup Q1a1a-M120, a paternal lineage connecting populations in Siberia and East Asia." *Annals of Human Biology* 46(3): 261-266.

Sun, Jin et al. 2021. "Shared paternal ancestry of Han, Tai-Kadai-speaking, and Austronesian-speaking populations as revealed by the high resolution phylogeny of O1a-M119 and distribution of its sub-lineages within China." *American Journal of Physical Anthropology* 174(4) 686-700.

Svoboda, Jiri et al. 2009. "Pavlov VI: An Upper Palaeolithic living unit." *Antiquity* 83: 282–295.

Tajima, Atsushi et al. 2004. "Genetic origins of the Ainu inferred from combined DNA analyses of maternal and paternal lineages." *Journal of Human Genetics* 49: 187-193.

Takakura, Jun 2012. "Emergence and development of the pressure microblade production: a view from the Upper Paleolithic of northern Japan." In: *The Emergence of Pressure Blade Making.* Edited by P. Desrosiers et al. Springer: Boston, MA. pp. 285-306.

Takakura, Jun 2020. "Rethinking the disappearance of microblade technology in the terminal Pleistocene of Hokkaido, Northern Japan: looking at archaeological and palaeoenvironmental Evidence." *Quaternary* 3: 21.

Takashi, Tsutsumi 2012. "MIS3 edge-ground axes and the arrival of the first Homo sapiens in the Japanese archipelago." *Quaternary International* 248: 70-78.

Tamang, Rakesh et al. 2018. "Reconstructing the demographic history of the Himalayan and adjoining populations." *Human Genetics* 137(2): 129-139.

Tambets, Kristina et al. 2004. "The western and eastern roots of the Saami - the story of genetic 'outliers' told by the mitochondrial DNA and Y Chromosome." *American Journal of Human Genetics* 74: 661-682.

Tamm, Erika et al. 2007. "Beringian standstill and spread of Native American founders." *Public Library of Science One* 2(9): e829.

Taylor, R.E., David Glenn Smith and John R Southon 2001. "The Kennewick skeleton: chronological and biomolecular contexts." *Radiocarbon* 43(2B): 965-976.

Taylor, William Treal Taylor et al. 2020. "Early pastoral economies and herding transitions in Eastern Eurasia." *Scientific Reports* 10: 1001.

Tcherenkov, Lev and Stéphane Laederich 2004. *The Rroma: otherwise known as Gypsies, Gitanos, Gyphtoi, Tsiganes, Ţigani, Çingene, Zigeuner, Bohémiens, Travellers, Fahrende, etc.* Schwabe Verlag.

Thangaraj, Kumarasamy et al. 2003. "Genetic affinities of the Andaman Islanders, a vanishing human population." *Current Biology* 13: 86-93.

Thomason, Sarah G. 2001. *Language Contact: An Introduction*. Washington, DC: Georgetown University Press.

Thurgood, Graham 1994. "Tai-Kadai and Austronesian: the nature of the historical relationship." *Oceanic Linguistics* 33(2): 345-368.

Tishkoff, Sarah A. et al. 2007. "History of click speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation." *Molecular Biology and Evolution* 24(10): 2180-2195.

Tofanelli, Sergio et al. 2009a. "On the origins and admixture of Malagasy: new evidence from high resolution analyses of paternal and maternal lineages." *Molecular Biology and Evolution* 26(9): 2109-2124.

Tofanelli, Sergio et al. 2009b. "J1-M267 Y lineage marks climate-driven pre-historical human displacements." *European Journal of Human Genetics* 17: 1520-1524.

Tranter, Nicholas 2012. "Introduction: typology and area in Japan and Korea." *The languages of Japan and Korea*. Edited by Nicolas Tranter. Oxon, UK: Routledge, pp. 3-23.

Trask, R.L. 1996. *Historical Linguistics*. London: Arnold.

Tremayne, Andrew 2015. "New Evidence for the timing of Arctic Small Tool Tradition coastal settlement in northwest Alaska." *Alaska Journal of Anthropology* 13(1): 1-18.

Triki-Fendri, Soumaya et al. 2015. "Paternal lineages in Libya inferred from Y-chromosome haplogroups." *American Journal of Physical Anthropology* 57: 242–251.

Trinkaus, Erik et al. 2003. "An early modern human from the Peştera cu Oase, Romania." *Proceedings of the National Academy of Sciences of the United States of America* 100(20): 11231–11236.

Trivedi, R. et al. 2008. "Genetic imprints of Pleistocene origins of Indian populations: a comprehensive phylogenetic sketch of Indian Y-chromosomes." *International Journal of Human Genetics* 8(1-2) 97-118.

Trombetta, Beniamino et al. 2015. "Phylogeographic refinement and large scale genotyping of human Y chromosome haplogroup E provide new insights into the dispersal of early pastoralists in the African continent." *Genome Biology and Evolution* 7(7): 1940–1950.

Tumonggor, Meryanne K. et al. 2014. "Isolation, contact and social behavior shaped genetic diversity in West Timor." *Journal of Human Genetics* 59: 494-503.

Turney, Chris S.M. and Heidi Brown 2007. "Catastrophic early Holocene sea level rise, human migration and the Neolithic transition in Europe." *Quaternary Science Reviews* 26: 2036–2041.

Underhill, Peter A. et al. 1996. "A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history." *Proceedings of the National Academy of Sciences of the United States of America* 93: 196-200.

Underhill, Peter A. et al. 2000. "Y chromosome variation and the history of human populations." *Nature Genetics* 26: 358-361.

Underhill, P.A. et al. 2001. “The phylogeography of Y chromosome binary haplogroups and the origins of modern human populations.” *Annals of Human Genetics* 65: 43-62.

Underhill, Peter A. et al. 2007. “New phylogenetic relationships for Y-chromosome haplogroup I.” In: *Rethinking the Human Revolution*. Edited by P. Mellars et al. Cambridge, UK: McDonald Institute for Archaeological Research, pp. 33-42.

Underhill, Peter A. et al. 2015. “The phylogenetic and geographic structure of Y-chromosome haplogroup R1a.” *European Journal of Human Genetics* 23: 124-131.

United Nations Educational, Scientific and Cultural Organization (UNESCO). Gobustan Rock Art Cultural Landscape. Accessed September 11, 2021. https://whc.unesco.org/en/list/1076/.

Vajda, Edward 2009. “The Languages of Siberia.” *Language and Linguistics Compass* 3(1): 424-440.

Valverde, Laura et al. 2016. “European paternal lineage M269: dissection of the Y-SNP S116 in Atlantic Europe and Iberia.” *European Journal of Human Genetics* 24: 437-441.

Varzari, Alexander et al. 2009. “Searching for the origins of Gagauzes: inferences from Y-chromosome analysis.” *American Journal of Human Biology* 21: 326-336.

Velichko, A.A. et al. 2009. “Initial human settlement of East European Plain.” In: *The East European Plain on the Eve of Agriculture*. Edited by Pavel M. Dolukhanov, Graeme R. Sarson and Anvar M. Shukurov. BAR International Series 1964. Oxford, UK: Archaeopress, pp. 17-21.

Venkateswar, Sita 1999. “The Andaman Islanders.” *Scientific American* 280 (5): 82-88.

Vennemann, Theo 1994. “Linguistic reconstruction in the context of European prehistory.” *Transactions of the Philological Society*, 92(2): 215-284.

Vennemann, Theo 2000. “Zur Entstehung des Germanischen.” *Sprachwissenschaft* 25(3): 233-269.

Völgyi, Antonia et al. 2008. “Haplogroup distribution of Hungarian population and the largest minority group.” *Forensic Science International Genetics Supplement Series* 1: 383-385.

Voskarides, Konstantinos et al. 2016. “Y-chromosome phylogeographic analysis of the Greek-Cypriot population reveals elements consistent with Neolithic and Bronze Age settlements.” *Investigative Genetics* 7:1.

Wang, Chuan Chao et al. 2013a. “Late Neolithic expansion of ancient Chinese revealed by Y chromosome haplogroup O3a1c-002611.” *Journal of Systematics and Evolution* 51(3): 280-286.

Wang, Chuan-Chao and Hui Li 2013b. “Inferring human history in East Asia from Y chromosomes.” *Investigative Genetics* 4:11.

Wang, Chuan-Chao et al. 2014. “Genetic structure of Qiangic populations residing in the Western Sichuan Corridor.” *Public Library of Science One* 9(8): e103772.

Waterman, John T. 1976. *A History of the German Language: with Special Reference to the Cultural and Social Forces that Shaped the Standard Literary Language.* Seattle: University of Washington Press.

Waters, Michael R. et al 2011. “Pre-Clovis mastodon hunting 13,800 years ago at the Manis Site Washington.” Science 334: 351-353.

Wedage, Oshan et al. 2019. “Specialized rainforest hunting by Homo sapiens ~45,000 years ago.” *Nature Communications* 10: 739.

Wei, Lan-Hai and Hui Li 2017. "Fuyan human of 120–80 kya cannot challenge the Out-of-Africa theory for modern human dispersal." *Science Bulletin* 62: 316-318.

Wei, Lan-Hai et al. 2017a. "Phylogeography of Y-chromosome haplogroup O3a2b2-N6 reveals patrilineal traces of Austronesian populations on the eastern coastal regions of Asia." *Public Library of Science One* 5:12(4): e0175080.

Wei, Lan-Hai et al. 2017b. "Phylogeny of Y-chromosome haplogroup C3b-F1756, an important paternal lineage in Altaic-speaking populations." *Journal of Human Genetics* 62: 915-918.

Wei, Lan-Hai et al. 2018. "Paternal origin of Paleo-Indians in Siberia: insights from Y-chromosome sequences." *European Journal of Human Genetics* 26: 1687-1696.

Wegwerth, Antje et al. 2021. "Black Sea hydroclimate and coupled hydrology was strongly controlled by high-latitude glacial climate dynamics." *Communications, Earth and Environment* 2: 63.

Westaway, K. et al. 2017. "An early modern human presence in Sumatra 73,000–63,000 years ago." *Nature* 548: 322–325.

White, Tim D. et al. 2003. "Pleistocene Homo sapiens from Middle Awash, Ethiopia." *Nature* 423: 742-747.

Whitman, John 2012. "The relationship between Japanese and Korean." *The languages of Japan and Korea*. Edited by Nicolas Tranter. Oxon, UK: Routledge, pp. 24-38.

Whitney, Bronwen et al. 2014. "Pre-Columbian raised-field agriculture and land use in the Bolivian Amazon." *The Holocene* 24(2): 231-241.

Wilkin, Shevan et al. 2020. "Dairy pastoralism sustained eastern Eurasian steppe populations for 5,000 years." *Nature Ecology and Evolution* 4: 346–355.

Wilkins, Jayne et al. 2020. "Innovative Homo sapiens behaviours 105,000 years ago in a wetter Kalahari." *Nature* 592: 248-252.

Williams, Jack F. 2003. "Who are the Taiwanese? Taiwan in the Chinese diaspora." In: *The Chinese Diaspora. Space, Place, Mobility and Identity*. Edited by Laurence J.C. Ma and Carolyn Cartier. Rowman and Littlefield Publishers, pp. 163-189.

Williams, Thomas J. and David B. Madsen 2020. "The Upper Paleolithic of the Americas." PaleoAmerica, 6(1): 4-22.

Wilson, Samuel M. 2007. The Archaeology of the Caribbean. Cambridge, UK: Cambridge University Press, pp. 59-136.

Winford, Donald 2003. *An Introduction to Contact Linguistics*. Malden, MA: Blackwell.

Winters, Clyde 2010. "Y-Chromosome evidence of an African origin of Dravidian agriculture." *International Journal of Genetics and Molecular Biology* 2(3): 30-33.

Wood, Elizabeth T. 2005. "Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biases demographic processes." *European Journal of Human Genetics* 13: 867-876.

Wu, Tianyi, and Bengt Kayser 2006. "High altitude adaptation in Tibetans." *High Altitude Medicine and Biology* 7(3): 193-208.

Wu, Qiao et al. 2020. "Phylogenetic analysis of the Y-chromosome haplogroup C2b-F1067, a dominant paternal lineage in Eastern Eurasia." *Journal of Human Genetics* 65: 823–829.

Xia, Zi-Yang et al. 2019. "Inland-coastal bifurcation of southern East Asians revealed by Hmong-Mien genomic history." bioRxiv preprint. https://doi.org/10.1101/730903.

Xue, Yali et al. 2006. "Male demography in East Asia: a north-south contrast in human population expansion times." *Genetics* 172: 2413-2439.

Yan, Shi et al. 2014. "Y chromosomes of 40% of Chinese descend from three Neolithic super-grandfathers." *Public Library of Science One* 9(8): e105691.

Yanchilina, Anastasia G. et al. 2017. "Compilation of geophysical, geochronological, and geochemical evidence indicates a rapid Mediterranean-derived submergence of the Black Sea's shelf and subsequent substantial salinification in the early Holocene." *Marine Geology* 383: 14-34.

Yang, Jian et al. 2017. "Genetic signatures of high-altitude adaptations in Tibetans." *Proceedings of the National Academy of Sciences of the United States of America* 114(16): 4189-4194.

Yang, Melinda A. et al. 2017. "40,000-year-old individual from Asia provides insight into early population structure in Eurasia." *Current Biology* 27: 3202-3208.

Yanko-Hombach, Valentina et al. 2007. "Controversy over the great flood hypotheses in the Black Sea in light of geological, paleontological, and archaeological evidence." *Quaternary International* 167–168: 91–113.

Yatoo, Mumtaz et al. 2020. "New evidence from the Kashmir Valley indicates the adoption of East and West Asian crops in the western Himalayas by 4400 years ago." *Quaternary Science Advances* (2): 100011.

YCC 2002. "A nomenclature system for the tree of human Y-chromosome binary haplogroups." *Genome Research* 12: 339-348.

Yi, Mingjie et al. 2015. "Rethinking the origin of microblade technology: A chronological and ecological perspective." *Quaternary International* 400: 130-139.

Yunusbayev, Bayazit et al. 2012. "The Caucasus as an asymmetric semipermeable barrier to ancient human migrations." *Molecular Biology and Evolution* 29(1): 359-365.

Zalloua, Pierre A. et al. 2008. "Y-chromosome diversity in Lebanon is structured by recent historical events." *The American Journal of Human Genetics* 82: 873-882.

Zegura, Stephen L. et al. 2004. "High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas." *Molecular Biology and Evolution* 21(1): 164-175.

Zerjal, Tatiana et al. 1997. "Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis." *American Journal of Human Genetics* 60: 1174-1183.

Zhang Chi and Hsiao-chun Hung 2008. "The Neolithic of Southern China – origin, development, and dispersal." *Asian Perspectives* 47(2): 299-329.

Zhang Chi and Hsiao-chun Hung 2010. "The emergence of agriculture in southern China." *Antiquity* 84: 11-25.

Zhang, Chi and Hsiao-chun Hung 2012. "Later hunter-gatherers in southern China, 18000–3000 BC." *Antiquity* 86: 11-29.

Zhang, Dong Ju et al. 2016. "History and possible mechanisms of prehistoric human migration to the Tibetan Plateau." *Science China Earth Science* 59(9): 1765-1778.

Zhang, Xiaoming et al. 2015. “Y-chromosome diversity suggests southern origin and Paleolithic backwave migration of Austro-Asiatic speakers from eastern Asia to the Indian subcontinent.” *Scientific Reports* 5:154 86.

Zhao, Zhijun 2011. “New archaeobotanic data for the study of the origins of agriculture in China.” *Current Anthropology* 52(S4): S295-S306.

Zhivotovsky, Lev. A. et al. 2004. “The effective mutation rate at Y-chromosome short tandem repeats, with application to human population-divergence time.” *American Journal of Human Genetics* 74: 50-61.

Zhong, Hua et al. 2011. “Extended Y-chromosome investigations suggests postglacial migrations of modern humans into East Asia via the northern route.” *Molecular Biology and Evolution* 28(1): 717-727.

Zvelebil, Marek 2006. “Mobility, contact, and exchange in the Baltic Sea basin 6000–2000 BC.” *Journal of Anthropological Archaeology* 25: 178–192.

Zwyns, Nicolas et al. 2019. “The northern route for human dispersal in central and northeast Asia: new evidence from the site of Tolbor-16, Mongolia.” *Scientific Reports* 9:11759.

About Dr. St. Clair

Dr. Michael St. Clair currently lives in Stuttgart, Germany with his wife and son. He holds a Ph.D. in German linguistics from the University of California, Berkeley. As an undergraduate at the University of Colorado, Dr. St. Clair attended introductory biology, chemistry, and anthropology courses to satisfy the core curriculum requirements. This provided a foundation for exploring the potential of genetic data as a tool for linguistic research. In 2012, he delivered a presentation at a linguistics conference in order to showcase data from his dissertation. He explained that Y-chromosome variation in Europe helps to decipher the prehistory of Germanic languages. Afterwards a member of the audience suggested that his argument would be more persuasive by demonstrating the usefulness of Y-chromosome data for non-Germanic languages. In 2013, he started work on the Genetic-Linguistic Interface project. The project had two goals. The first goal was to acquire sufficient data. A database was ultimately constructed to manage the large volume of information uncovered by his research. The second goal was to develop a methodology for employing genetic data as a tool for linguistic research. Dr. St. Clair published an open access monograph in December of 2021 that explains his methodological solution.